AFRL-VS-HA-TR-2006-1052

Evaluation of Cross-Correlation Methods on a Massive Scale for Accurate Relocation of Seismic Events in East Asia

P.G. Richards
F. Waldhauser
W.Y. Kim
D. Schaff

Lamont-Doherty Earth Observatory of Columbia University
61 Route 9W
Palisades, NY 10964

21 April 2006

Final Report

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
29 Randolph Road
AIR FORCE MATERIEL COMMAND
Hanscom AFB, MA 01731-3010

This technical report has been reviewed and is approved for publication.

AFRL-VS-HA-TR-2006-1052

/signed/

ROBERT J. RAISTRICK
Contract Manager

/signed/

ROBERT BELAND, Chief
Battlespace Surveillance Innovation Center

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the National Technical Information Service.

If your address has changed, if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/VSIM, 29 Randolph Rd., Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
21-04-2006	Final Technical Report	01-10-2003 - 07-01-2006

4. TITLE AND SUBTITLE
Evaluation of Cross-Correlation Methods on a Massive Scale for Accurate Relocation of Seismic Events in East Asia

5a. CONTRACT NUMBER
F19628-03-C-0129

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62601F

6. AUTHOR(S)
Paul G. Richards, Felix Waldhauser, David Schaff, and Won-Young Kim

5d. PROJECT NUMBER
1010

5e. TASK NUMBER
SM

5f. WORK UNIT NUMBER
A1

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lamont-Doherty Earth Observatory of Columbia University
61 Route 9W
Palisades, NY 10964

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
29 Randolph Road
Hanscom AFB, MA 01731

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/VSBYE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-VS-HA-TR-2006-1052

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
A weakness of traditional methods of seismic event location is their susceptibility to errors in picking arrival times of regional phases such as *Pn, Pg,* and teleseismic phases. We evaluate the practical utility of waveform cross-correlation as a way to reduce or remove pick error, applied to large datasets in seismically active regions of East Asia, for purposes of obtaining significant improvements in event location.

Key ideas in this project emerged during 2000-2003 from work of a Lamont-led consortium to calibrate stations in East Asia. Thus, we found for a number of regions that one of the best methods for obtaining ground truth events was to use waveform cross-correlations that enabled excellent (sub-kilometer) precision in relative locations. Additional (often, non-seismic) information then allowed absolute locations to be estimated. A preliminary study of China has shown that 1300 out of 14,000 earthquakes (approximately 9%) exhibit high cross-correlations with at least one other earthquake, and on this basis we have found 494 sets of cross-correlated multiplets, ranging from doublets to one multiplet with 26 events.

We apply waveform cross-correlations to the problem of event location in four project areas, namely China, parts of Eastern Canada, the Central United States (New Madrid), and (for purposes of validation of our overall method) California. For each project area, we carry out the three steps of data acquisition (digital waveforms, traditional phase picks, catalog information), waveform cross-correlations (exploring the effects of different filters, spatial event separation, and differences in magnitude), and event location (double-difference).

We produce datasets of waveforms, cross-correlation measurements of arrival differences, and sets of precise locations. However, we regard our main work in this project to be an evaluation of a potentially powerful method—one that can be expected to become increasingly important as the number of digitally well-documented events in a particular region increases, thus increasing the probability that waveform correlation methods can be applied on a large scale.

15. SUBJECT TERMS
Seismology, seismic monitoring, earthquake location

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Robert J. Raistrick
Unclassified	Unclassified	Unclassified	Unlimited	75	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

Summary	1
Section One: Reference Events for Regional Seismic Phases at IMS Stations in China	2
Abstract	2
Introduction	3
ABCE Data and Double-Difference Cluster Analysis	4
Generation of Reference Events	8
Evaluation of Solution Quality	11
Conclusions	15
Acknowledgments	16
Section Two: Double-Difference Relocation of Earthquakes in Central-Western China	33
Section Three: The Applicability of Modern Methods of Earthquake Location	36
Abstract	36
Introduction	37
Background	38
Method of Analysis for the Different Regions	43
Preliminary Results for Four Different Regions	45
Discussion and Conclusions	49
Acknowledgments	53

SUMMARY

A weakness of traditional methods of seismic event location is their susceptibility to errors in picking arrival times of regional phases such as *Pn, Pg,* and teleseismic phases. We evaluate the practical utility of waveform cross-correlation as a way to reduce or remove pick error, applied to large datasets in seismically active regions of East Asia, for purposes of obtaining significant improvements in event location.

Key ideas in this project emerged during 2000-2003 from work of a Lamont-led consortium to calibrate stations in East Asia. Thus, we found for a number of regions that one of the best methods for obtaining ground truth events was to use waveform cross-correlations that enabled excellent (sub-kilometer) precision in relative locations. Additional (often, non-seismic) information then allowed absolute locations to be estimated. A preliminary study of China has shown that 1300 out of 14,000 earthquakes (approximately 9%) exhibit high cross-correlations with at least one other earthquake, and on this basis we have found 494 sets of cross-correlated multiplets, ranging from doublets to one multiplet with 26 events.

We apply waveform cross-correlations to the problem of event location in four project areas, namely China, parts of Eastern Canada, the Central United States (New Madrid), and (for purposes of validation of our overall method) California. For each project area, we carry out the three steps of data acquisition (digital waveforms, traditional phase picks, catalog information), waveform cross-correlations (exploring the effects of different filters, spatial event separation, and differences in magnitude), and event location (double-difference).

We produce datasets of waveforms, cross-correlation measurements of arrival differences, and sets of precise locations. However, we regard our main work in this project to be an evaluation of a potentially powerful method—one that can be expected to become increasingly important as the number of digitally well-documented events in a particular region increases, thus increasing the probability that waveform correlation methods can be applied on a large scale.

SECTION ONE

Reference Events for Regional Seismic Phases at IMS Stations in China

Felix Waldhauser and Paul G. Richards[1]

Lamont-Doherty Earth Observatory, Columbia University

P.O. Box 1000, Palisades, NY 10964

felixw@ldeo.columbia.edu

[1)]also, Department of Earth and Environmental Sciences

Abstract Seismic event location within the context of monitoring the Comprehensive Nuclear-Test-Ban Treaty (CTBT) entails *a priori* knowledge of the travel time of seismic phases for a given source to stations of the International Monitoring System (IMS). Such travel time information (or ground truth, GT) is provided empirically by seismic reference events, events that have well-determined hypocenter locations (epicenters typically known to +/- 5 km with high confidence) and origin times. In this study we present new reference events for the calibration of six seismic stations of the IMS in China, a region with high seismic activity. We use the Annual Bulletin of Chinese Earthquakes, which lists about 1,000 earthquakes in and near China each year with consistent phase picks at regional stations, to determine precise relative earthquake locations from double-difference cluster analysis. The resulting high-resolution image of active faulting at seismogenic depths in areas of dense seismicity is correlated with the tectonic structure derived from mapped fault information at the surface to validate the absolute locations. We generated 59 reference events with $M \geq 3.5$, distributed in six clusters in central and eastern China, and recorded by at least one of the six IMS stations. The scatter in relative travel time residuals is reduced from 1.28 s before to 0.61 s after relocation, consistent with the relocated position of the events. The degree of correlation between seismicity structure and well-characterized fault data indicates that, in four clusters, the locations of the new reference events are accurate to within 5 km (GT5), and in two clusters within 10 km (GT10).

Introduction

Effective monitoring of compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT) requires prompt and accurate characterization of about 100 seismic events per day. Such characterization entails the need for accurate location estimates, which in turn requires knowledge of the travel time of seismic phases such as *Pg*, *Pn*, teleseismic *P*, and their *S*-wave analogs for a given source-station configuration. It has become conventional to describe these travel times in terms of Source Specific Station Corrections (SSSCs), which are added to the travel times predicted by a standard travel time model (usually taken as IASP91, see Kennett and Engdahl, 1991) to obtain travel times at a particular station as a function of distance, azimuth, and depth. When implemented at the International Data Centre (IDC) in Vienna, for stations of the International Monitoring System (IMS), SSSCs are expected to improve event locations by removing location bias due to unmodeled velocity structure between source and receiver. With the increasing amount and quality of seismic data collected by the IMS, unmodeled velocity structure remains the main cause of significant errors and uncertainties in the location of seismic events for monitoring purposes. This is especially true for regional signals, whose travel time can be significantly different (fast or slow) compared to the predictions of travel time models that represent global averages.

In recent years, SSSCs have been developed using seismic velocity models of the crust and upper mantle for the region surrounding an IMS station (e.g., Yang et al., 2001; Murphy et al., 2002; Ritzwoller et al., 2003). These models are generally based on active or passive seismic data, or a combination of both, and travel times have been computed through these models from a station to a set of surface grid points within 20° distance. A critical step in producing SSSCs is the process by which these correction surfaces are validated against independent data, such as reference events. Reference events are seismic events whose location and origin time are known independent of the monitoring network. Their location uncertainty within the 90% confidence level is generally referred to as the ground truth level of an event, GT*x*, where *x* specifies the epicenter location accuracy in kilometer (i.e., true epicenter lies within *x* kilometer of the estimated epicenter) (Bondár, 2001). GT0 reference events, for example, have epicenters and origin times known to within 100 m and 0.1 sec, respectively, and are typically obtained from peaceful nuclear or chemical explosions (e.g., Sultanov et al., 1999). Since the global distribution of man-made sources with well known source parameters is sparse, data from moderate size earthquakes must

also be used. The location of earthquakes, however, is generally much less well known. Typically events of GT5 quality or better are obtained when the earthquake occurred within a local network. A significant effort is necessary to turn earthquake locations into reference events of GT5 quality when they occur within regional networks.

In this study we follow an integrated approach to generating reference events from regional network data for IMS station calibration in central and eastern China. The complex tectonics in this region are expressed by quite high but diffuse seismicity. We analyze more than 11,000 events in the Annual Bulletin of Chinese Earthquakes (ABCE) from 1985 to 1999 for their potential use as reference events. Many of these events are recorded at ABCE stations which are close to or co-locate with the planned IMS stations (Table 1). We determine 59 reference events at the GT5 and GT10 level in central and eastern China by combining precise relative event relocations obtained from cluster analysis of ABCE data with a data base of mapped fault information at the surface. While the motivation of this work was within the context of CTBT monitoring, the approach outlined here may be useful for other purposes, for example to image faults at seismogenic depths over large areas for seismic hazard investigations. For such studies, however, the ABCE data needs to be combined with provincial and local network data across China.

ABCE Data and Double-difference Cluster Analysis

An electronic version of the Annual Bulletin of Chinese Earthquakes (ABCE) (z-files, included in the IASPEI International Handbook of Earthquake and Engineering, Part B, 2003) includes about 14,000 events in and near China with magnitudes up to M 6.8 which occurred between 1985-1999, with more than 10,000 events located in mainland China (Figure 1). P-and S-phase picks are available for 170 regional stations in and near China. These picks are remarkably consistent as noted by Hearn and Ni (2001) and demonstrated by the analysis of travel time residuals below. About 36,000 P-phase picks, selected from 11,500 events recorded at the ABCE stations BJI, HLR, KMI, LZH, SSE, and XAN and located within 20° of each station, are analyzed for their accuracy. These six stations are at or close to the designated sites of IMS stations in China (Table 1). 96% of the ABCE Pn-phase picks are reported to a tenth of a second (Figure 2a). Locations of 43% of the events are rounded to the nearest tenth of a degree, introducing (presum-

ably randomly distributed) location uncertainties of up to 5 km (Figure 2b). Depths for about 50% of the events appear to be fixed at 10, 15, or 33 km (Figure 2c).

In addition to an electronic ABCE we have access to a printed version of the ABCE for events that occurred in 1985 and 1986, and in the years 1991-1995. The printed ABCE reports earthquake locations to the nearest hundredth of a degree, and they appear to be a revised version of the electronic ABCE. Comparison between event locations on mainland China in the printed ABCE and the corresponding events in the electronic ABCE indicates a mean event mislocation of about 11 km in both horizontal directions. Comparison of the electronic ABCE with data from a local network in the Sichuan/Yunnan Province (Z. Yang, personal communication) indicates similar differences, but individual errors may be larger in some areas.

Due to uncertainties of this order in ABCE locations we can not extract high-quality reference events directly from the catalog, but accurate event locations may be determined by relocating the events. Relocation of the ABCE data for the purpose of obtaining reference events is problematic, however, because of the sparse distribution of available stations for most of the events. None of the events in the ABCE have the potential to achieve GT5 status at the 95% confidence level based on the seismic network criteria put forward by Bondár et al. (2004). These criteria require at least 10 stations within 250 km with an azimuthal gap less than 110°, a secondary azimuthal gap less than 160°, and at least one station within 30 km from the epicenter. We can turn instead to relative location methods such as the double-difference technique of Waldhauser and Ellsworth (2000) to reduce the effect of model errors, and then use near surface information to constrain the absolute locations.

The fundamental equation of the double-difference algorithm relates the differences between the observed and predicted phase travel time difference for pairs of earthquakes observed at common stations to changes in the vector connecting their hypocenters. By choosing only relative phase travel times for events that are close together (i.e., closer than the scale-length of the surrounding velocity heterogeneity), wave paths outside the source region are similar enough so that common mode travel time errors are canceled for each linked pair of events. It is then possible to obtain high-resolution hypocenter locations over large areas without the use of station corrections. This approach is extremely useful in our search for potential reference events across China, as it allows us to efficiently relocate dense seismicity across large areas. In some of the clusters investigated in this study the seismicity spreads over more than a hundred km distance, in

which case we linked events over short distances (typically smaller than 10 km) to build a chain of links that connect events across the entire cluster. Other multiple-event location algorithms such as JHD (Douglas, 1967) or HDC (Jordan and Sverdrup, 1981) employ station corrections that are fixed for a particular cluster of events, thus limiting the spatial area within which events can be relocated.

The presence of severely mislocated events in the ABCE can hamper the inversion, since the linearization of the non-linear double-difference location problem (which solves for adjustments to initial locations) requires the initial locations to be close to their true value. To stabilize the inversion we search for clusters of well linked event pairs (at least 10 stations), and iteratively solve the system of weighted double-difference equations by means of least squares using the program *hypoDD* (Waldhauser, 2001). Convergence to a stable solution is greatly helped by the high quality of the ABCE phase picks, which have little contamination by outliers. A search of the ABCE for clusters of events that are most suitable for double-difference relocation in terms of network geometry and event density resulted initially in 36 clusters, ranging in size between 20 and 344 earthquakes. This analysis was done by searching about 460,000 *P*- and *S*-phase picks. For the 36 clusters, *P*- and *S*-phase pairs at common stations out to 2000 km distance were formed. Regional 1-D layered velocity models were used to predict the travel time differences and partial derivatives. These models were adapted from the models used for routine locations at Chinese provincial seismographic networks (Jih, 1998).

In many of the 36 clusters the high resolution relocations reveal detailed structural information about the active fault on which they occur, such as dip and strike. It is possible to validate the absolute location accuracy of such an event cluster by comparing the relocated seismicity with independent surface information. Extremely detailed fault data for mainland China have been prepared by the U.S. Geological Survey (USGS Astrogeology Team, lead by Philip A. Davis), derived from volumes on the regional geology of Chinese provinces published by the Geological Publishing House (Beijing, 1984-1993). The GIS data base includes all fault lines, and additional fault parameters such as type of fault, age and name of fault, relative depth of the fault, how the existence of the fault was inferred, and the direction and amount of dip on the fault plane. These additional parameters, however, are not reported for all faults. Also, the location of the fault lines are subject to uncertainties due to inconsistencies in the published maps and/or uncertainties related to the digitizing and GIS mapping process, according to the unpublished database docu-

mentation 'Compilation of bedrock geology and structure databases of China, including relevant ancillary databases compiled by Los Alamos National Laboratory'. In general the uncertainties range from a few hundreds of meters to a few kilometers in rare cases. Specifically, the information on deep faults used in this study were drawn with very thick lines on the published Chinese maps, which translate to a possible maximum uncertainty of 0.5 km in the digital map provided by the USGS.

In order to extract potential reference events for the purpose of IMS station calibration from any of the 36 clusters, we have defined the following three criteria: The relocated seismicity

- includes events of $M \geq 3.5$ that have small relative location errors and are recorded at at least one of the planned IMS stations (or a surrogate) within 2000 km distance,
- indicates fault structure such as strike and dip,
- and correlates with nearby major deep reaching faults or faulting patterns included in the Chinese fault data base.

The 36 clusters are reviewed in terms of their location close to faults labeled as 'deep reaching' in the data base, or faults that have additional information to support their association with the seismic activity. In most of the 36 cases, no unique association with mapped surface traces is possible. Figure 3 shows two of such clusters in Guizhou Province and Sichuan Province. For the Guizhou cluster (Figure 3a) the seismicity spreads over several faults, indicating that the events do not occur on a single fault, but rather on different adjacent faults. Only a few events appear to occur on a deep reaching, right lateral strike-slip fault (thick line in Figure 3a), making a clear association difficult. Figure 3b shows a cluster of events that locates within an aseismic block bounded by deep-reaching faults to the southwest and the southeast, more than 40 km from any mapped surface trace. While it is quite possible that these locations are accurate to within a few kilometers, we can not verify that accuracy with independent surface data, which our approach to build reference events requires.

Our detailed investigation of the 36 clusters resulted in a subset of six that has a positive correlation between the high-resolution seismic locations and available well-characterized fault data. These six clusters include a total of 262 events, located in central and eastern China in the provinces of Sichuan (near the cities of Neijiang and Batang), Shanxi (near Datong), Tibet (near Tangmai), Qinghai (near Menyuan), and Yunnan (near Jinggu) (Figure 1). All six clusters include events with $M \geq 3.5$. (For the events we used whose magnitudes were not reported in the ABCE

catalog, the large distances out to which these events were recorded indicate that they have magnitudes well above 3.5.)

Generation of Reference Events

Events in the Neijiang cluster are used to demonstrate the improvement in event relocations over the ABCE locations. This cluster in Sichuan Province includes 61 events that were recorded at 31 stations between 1989-1999 (Figure 4). The maximum separation between events that are linked together by common phase pairs is 10 km, while the cluster dimension is about 25 km. The largest azimuthal station gap is 40° and the closest station is 120 km away (Table 2). Table 3 lists the regional (Sichuan/Yunnan) velocity model used for relocation. Figure 5 compares the double-difference epicenter locations with the locations listed in the ABCE. The double-difference locations indicate a much tighter distribution compared to the ABCE, and concentrate near a deep north dipping reverse fault mapped at the surface by Chinese scientists (thick fault trace in Figure 5). Shifts between ABCE and double-difference locations range from 0.5 to 40 km, with a mean of about 13 km. Shifts in origin times range from -4.9 to 4.3 s, with a mean of -0.08 s and a standard deviation of about 2 s. The root mean square (*rms*) residual after relocation is 0.7 s, down 72% from the initial value. A bootstrap analysis of the remaining differential time residuals (see Waldhauser and Ellsworth, 2000, for details) has been performed to assess the relative location uncertainty at the 90% confidence level. The *rms* values for semi-major and minor axes are 1.5 km and 0.85 km, respectively, and 1.5 km for errors in depths, for all 61 events.

Figures 6a-f show the relocation results for the six clusters, both in map view and fault perpendicular cross section (along A-A'), superimposed on the near surface information from the Chinese fault data base. Relative location errors from the bootstrap analysis are indicated in map view as ellipses and in the fault-perpendicular cross section as crosses. Station distributions for each cluster are indicated in Table 2, velocity models used for relocation in Table 3, and hypocentral parameters for the selected reference events in Table 4. Fault traces in map views of Figure 6 are represented by lines, with thick lines indicating that they are deep reaching faults. Additional structural information is included when available. Fault traces in cross section of Figure 6 are represented by triangles and are the projection of the main fault onto the cross section.

Although the double-difference algorithm is somewhat sensitive to absolute locations through a chosen model (see Waldhauser and Ellsworth, 2000), we have fixed, for all six clusters, the absolute centroid location (epicenter and depth) of the double-difference solutions at the position of the centroid derived from the events in the electronic ABCE or the printed ABCE. A comparison between the two catalogs for events in 1985 and 1986, and in the years 1991-1995 (years for which the printed ABCE is available), shows that the difference in centroid location is less than 4 km for the clusters near Neijiang, Datong, Batang, and Tangmai. For the Neijiang cluster, for example, it is about 3 km for the 25 events that are included in both catalogs (see stars in Figure 5a). Even though it appears that the printed ABCE is a revised version of the electronic ABCE, we use the centroid location from the events in the electronic ABCE for the 4 clusters because the electronic ABCE includes all events compared to the printed ABCE with limited number of years. Furthermore, the differences in hypocentroid locations are smaller than the GT5 level we aim to achieve. For the two clusters near Menyuan and Jinggu, however, the differences in cluster centroids derived from the electronic and the printed ABCE are larger than 5 km. The Menyuan cluster centroid as taken from the printed ABCE is about 0.05 °to the north and 0.035 ° to the east from the one computed from the electronic ABCE. For the Jinggu cluster the shift is 0.01 °and 0.09 °to the south and east, respectively. For these two clusters we use the centroids from the printed ABCE, because of the significant deviation in centroid locations and the fact that most events in these two clusters are included in the printed ABCE. Note that in none of the six cases did we move the cluster of relocated events to line up with the surface trace of the fault. The fault information is considered independent data which is used to validate the absolute location of these clusters of reference events (as taken from the ABCE bulletins), and to investigate their level of accuracy, in a tectonic context, as discussed in the following.

The relocated seismicity in the Neijiang cluster (Figure 6a) images a ~25 km deep, about 10° northwest dipping fault which correlates well with an isolated mapped surface trace described as a deep reverse and north-dipping fault in the fault data base (thick line in map view in Figure 6a). In cross section, the projection of this fault along A-A' is indicated by a solid triangle. Eight events with $M \geq 4$ locate on this deep reverse fault, for which the inferred downward projection is indicated by a dashed line. Given the relative location uncertainty, the width of the fault zone imaged by these eight events is not resolvably different from zero. A ninth $M \geq 4$ event, having larger uncertainty, locates to the southeast away from the fault. Considering the formal uncer-

tainty in relative locations being smaller than 2 km for the eight M ≥ 4 events, and their correlation with independently mapped surface information, we select these eight larger events as reference events accurate to within 5 km (solid dots in cross section in Figure 6) (see Table 4 for hypocentral parameters). The position of the events along the strike of the fault is similar to the locations reported in the printed ABCE for events included in both bulletins (see Figures 5b). They are also similar to location results for 20 events derived from a simultaneous inversion of local network data (Z. Yang, personal communication) for 1-D velocity structure and earthquake locations (program VELEST, Roecker, 1977; Kissling et al., 1994).

Figure 6b shows results for the cluster near Datong. This cluster of 42 events was relocated using a velocity model for the region of NE China (Table 3). Most of the events occurred during a swarm-like activity in 1989, including 16 of the largest events (based on the number of stations that recorded them). In map view, the cluster of events lines up with a northeast trending deep fault, southeast and perpendicular to which are a series of northwest trending normal faults. Parallel and about 10 km to the northwest of the main fault exists what seems to be a secondary, deep fault. Even though this secondary fault is only 10 km away from the cluster centroid, the complex distribution of the relocated events (no narrow fault zone is imaged) and the indication of seismicity occurring on the normal faults to the southeast leads us to the conclusion that the 16 reference events are associated with the main fault, and stress is transferred to the normal faults that subsequently fail in smaller events.

The 68 events near Batang in Sichuan Province (Figure 6c) were relocated using the Sichuan velocity model (Table 3). The closest station, BTA, is only about 20 km away (see triangle in map view of Figure 6c), but no phase picks are listed in the ABCE bulletin for the events investigated. It is possible, though, that phase picks from this station were used in determining the locations reported in either the electronic or the printed ABCE, but for some reason are not reported in the bulletins. The relocated events mainly cluster around a NNW striking fault trace that is described as a west dipping reverse fault. This fault is paralleled by two other reverse faults at some 10 km distance to the southwest and northeast. The reverse faults intersect a left-lateral strike slip fault to the northwest. None of the faults are labeled as deep, even though the seismicity reaches depths of 25 km. The relocated seismicity correlates well with the fault information at the surface, imaging a WSW dipping fault with an effective width less than 1 km. The 7 reference events have magnitudes between 5.1 and 6.2, and are followed by aftershocks, some of which

appear to have occurred on the adjacent reverse faults to the northeast and southwest.

Near Tangmai in eastern Tibet 29 events were relocated using the velocity model for the Tibetian Plateau shown in Table 3. The events locate between two WNW-striking thrust faults (Figure 6d). They image a NNE dipping 35 km deep structure, a depth not unusual in this part of China. The centroid of the epicenter is within 10 km distance from the thrust fault to the south, but the downward projection of the surface trace is less clear because of the somewhat large horizontal relative errors for the 16 selected reference events, and missing information on the dip of the fault. Thus we place an upper bound of 10 km to the absolute location uncertainty of these events.

In the Menyuan cluster (Figure 6e) the relocated seismic activity correlates with a deep northwest striking reverse fault, right where a small left lateral stepover configuration exists (see inset in Figure 6e). We argue that this fault irregularity is the cause of the somewhat complex faulting pattern observed at depth, and thus can be used as a benchmark to constrain the absolute location of this cluster. A total of 12 events with $M > 4$ are selected, five of which are well constrained and are selected as reference events (Table 4). Additional seismic activity correlates with an adjacent, parallel fault to the southwest, and a nearby thrust fault to the northwest of the reference events.

In the Jinggu cluster (Figure 6f), most of the relocated seismicity appears to occur on a deep fault and possibly on an adjacent, parallel fault to the west, including 6 of the larger events selected as reference events. The faults are not further specified in the fault data base, but the relocated seismicity indicates that they are dipping to the west. A few isolated events seem to occur on additional faults, and some are associated with synclines. While there is a general agreement between the seismicity and the surface information, the absolute location of the reference events within the tectonic framework given by the limited fault information available to us is not as well constrained as in other clusters. Furthermore, relative location errors are larger than in most other clusters. We therefore consider these reference events to be within 10 km of the true locations.

Evaluation of Solution Quality

Several tests were performed to evaluate the stability of the double-difference solutions. In addition to the regional models listed in Table 3, a standard crustal model is used (IASP91) that

consists of a 20 km thick upper crust with a P-wave velocity of 5.8 km/s on top of a 15 km thick lower crust (vp = 6.5 km/s), and a upper mantle velocity of vp = 8.04 km/s. Figure 7 compares, for each cluster, the locations of reference events obtained with the regional model (thick black ellipses representing errors at the 90% confidence level, identical to final locations as shown in Figure 6) with locations from the IASP91 model (thin black ellipses). Note that even though we only show the reference events in Figure 7, all events in a particular cluster are used in the relocation procedure. In general, differences are less than 1 km between the two locations, and error ellipses for locations derived from the IASP91 velocity model are larger than those obtained from the regional models. Error ellipses overlap for corresponding events in almost all cases.

A second test used a jackknife method to investigate the influence of each reference event on the location of all others. As each event is linked to others through direct measurements, a less well constrained event may affect the relative location between others. Each cluster, therefore, was repeatedly relocated with one reference event removed at a time. The results of NEV*(NEV-1) locations are displayed as gray ellipses in Figure 7. The mean in horizontal and vertical deviation of each test location from the event's final location (solid ellipse) is 190 m and 340 m. Error ellipses are similar in most cases.

A third test, again using a jackknife method, assessed the effect of variation in station distribution on event locations. It involved removing one station at a time, each time locating all events within a cluster using the remaining NSTA - 1 stations. The results of NSTA * NEV locations are displayed as dots in each subfigure of Figure 7. The mean deviation of each test location from the centroid of all test locations for a particular event is 200 m in horizontal and 360 m in vertical direction. 95% of the 2881 samples are contained within an ellipse that has a major and minor axis of 1.5 km and 1 km, respectively. The results from these three tests indicate that changes in the model used to relocate the events, effects from individual reference events on others, and varying station distribution cause changes in relative locations that are generally less than 1 km. In each case the relocated reference events support the characterization of the corresponding faults as indicate in the Chinese fault data base.

For some events in the Tangmai and Jinggu clusters, relative depth errors are smaller than errors in the horizontal directions. The somewhat large horizontal errors are caused by azimuthal gaps in nearby stations (Table 2), while relative depths are still well constrained by downgoing ray paths of *Pn* phases observed at greater distances. Not that reference events in these two clusters

are less well constrained than in the other clusters.

Figure 8 shows absolute travel time residuals of the 59 reference events observed at the 6 IMS stations relative to the median travel time, for each cluster-station pair, and for double-difference (solid circles) and ABCE (open circles) solutions. Figure 9 indicates the IMS stations which recorded at least one reference event within a particular cluster. A significant reduction in residual scatter is observed, with the standard deviation decreasing from 1.28 s before to 0.61 s after relocation. The eight apparent phase pick outliers (solid squares in Figure 8) in the Neijiang, Datong, Menyuan, and Jinggu cluster may indicate timing problems. Note that such outliers, determined separately within each cluster, are generally downweighted or removed during the relocation process, and stations other than the affected IMS stations were used to relocate the events. The spread of residuals for events in the Tangmai cluster is larger than in other clusters, with the largest residual being larger than the resiudals of outliers identified in other clusters. This is likely because of the deviation of the true crustal structure from the IASP91 model in the source area, a region of thick lithosphere. The effect of model error is amplified by the large depth extent of the seismic activity to which *Pn* differential travel times are sensitive.

Since we keep the absolute location of the cluster centroid fixed at the initial locations, we are not able to quantify absolute location errors using standard approaches (e.g., by analyzing absolute travel time residuals), but instead validate the absolute location of the clusters and the reference events they include in a tectonic context derived from independent surface information. For comparison of uncertainty estimates based on absolute travel time residuals relative to the IASP91 travel times, we apply the location procedure *LocSat* (Bratt and Bache, 1988), currently in operation for nuclear test monitoring at the IDC in Vienna, to the ABCE arrival times of the 8 reference events near Neijiang. The resulting errors at the 95% confidence level have major axes ranging from 2.5 - 7.7 km. These major axes are predominantly oriented in the northwest direction perpendicular to the fault, a direction that is well constrained by the combined analysis of relocated seismicity and fault information. Thus we know the absolute locations of these events better then what is possible from the analysis of travel time residuals alone.

The promotion of seismic reference events to GT5 status (or any other GTn) is generally based on some quantitative measure of the absolute location uncertainty of an event. Bondár et al. (2004) proposed, after a thorough investigation of the effect of network geometry on the solution quality of earthquake and explosion locations, that such measures include the characteristics of

seismic networks used to locate a particular event. Clearly, the approach followed in this study lacks this type of quantitative measure, because none of the reference events derived here fulfills the GT5 network criteria of Bondár et al. (2004), at least not with the stations available to us. On the other hand, one can imagine the case where ground truth information is available from sources other than seismic data (i.e., surface rupture), providing a reference event at the GT1 level or better, even though stations that might have recorded the event would not meet the particular criteria proposed by Bondár et al. (2004).

There is a subjective component associated with our procedure in that the relocated seismicity is correlated with (independent) fault information to validate the absolute locations derived from the Chinese bulletins. This makes it difficult to quantitatively estimate the absolute location uncertainty from which to derive the GT level of each reference event. While errors in the relative location of events within each cluster, and errors in the mapping accuracy of the fault traces, are shown to be generally less than 1 km, errors due to fault mis-association can be larger. To estimate upper bounds for these errors we probe distances out to which a given cluster can be moved without jeopardizing the correlation between seismicity and fault information. For the clusters near Neijiang, Datong, Batang, Menyuan, the relocated seismicity we find that they can not be moved by more than 5 km without causing significant disagreement between the seismicity at depth and the active faulting pattern observed at the surface (Figure 6 and discussion above). For the reference events in the Neijiang cluster, location results using data from a local seismic network limit movement along the strike of the fault to significantly less than 5 km. For the reference events in these four clusters, therefore, we claim that they are accurate to within 5 km, promoting them to GT5 status (Table 4). The reference events in the Tangmai and Jinggu clusters are less accurately determined, because of the lower resolution in relative locations, and the range of possible correlation with the available fault information. We consider these events accurate to within 10 km, indicating solution qualities at the GT10 level (Table 4). The underlying assumption in assigning these GT levels is that the relocated seismicity actually occurs on the fault we use to validate the absolute locations with. In the case of the six clusters of reference events presented here, however, no other mapped fault near the reference events (and within the approximate absolute location uncertainty of the ABCE locations) is likely to be able to accommodate several earthquakes of M > 4.

Conclusions

The Annual Bulletin of Chinese Earthquakes is used to relocate events in six clusters in Central and Eastern China to image in detail the active fault at seismogenic depths. The relocated seismicity is associated to fault traces mapped at the surface and related structural information to validate the absolute location of each cluster as derived from the Chinese bulletins. 59 seismic reference events suitable for calibration of IMS stations in China (and possibly in nearby regions also) are then selected based on event magnitude, relative location errors, and consistency with the fault information. Substantial reduction in scatter of relative travel time residuals within each cluster of reference events is achieved, consistent with the relocated position of these events. Reference events determined in four of the six clusters have solution quality at the GT5 level, based on the degree of correlation between seismicity and faulting information. Reference events in the remaining two clusters are GT10. Each reference event provides P- and, in some cases, S-phase travel time information to at least one of six operational or planned IMS stations (Figure 9).

Prior to this study, reference events for the calibration of IMS stations in China have only been derived from nuclear explosion data, for regions in western China (Fisk, 2002; Waldhauser et al., 2004). Active tectonism from the ongoing collision between the Indian and the Asian continent, however, generates several tens of thousands of earthquakes every year in and near China. Thousands of seismic stations operated by local, provincial and regional networks record these earthquakes down to low magnitudes. Combining these data across boundaries of individual networks and provinces would substantially increase the density of recorded seismicity across China. The approach outlined here may then be applied on a much larger scale, to obtain accurate event locations for entire fault systems, in combination with good information on surface faulting. With about 190 $M \geq 4$ events listed in the ABCE per year, the number of reference events could then be substantially increased. Additionally, accurately located reference events enhance tomographic studies and can improve quantification of seismic hazard.

Acknowledgments

We are indebted to Chinese seismologists, without whose dedicated work in creating the ABCE this work could not have been carried out. We thank Tom Hearn for his advice on working with ABCE data, Bill Leith for providing us with a copy of the Chinese digital fault map, and Greg Yetman and Francesca Pozzi for their help with reading the GIS format. Reviews by Bob Engdahl and an anonymous reviewer greatly helped to improve the manuscript. This study was supported by contract DTRA 01-00-C-0031 of the Defense Threat Reduction Agency and by contract F19628-03-C-0129 of the Air Force Research Laboratory. This is Lamont-Doherty Observatory Contribution number 6666.

References

Bondár, I., X. Yang, R.G. North, and C. Romney (2001). Location Calibration Data for CTBT Monitoring at the Prototype International Data Center, *Pure appl. Geophys.*, **158**, 19-34.

Bondár, I., S.C. Myers, E.R. Engdahl, and E.A. Bergman (2004). Epicenter accuracy based on seismic network criteria, *Geophys. J. Int.*, **156**, 483-496.

Bratt, S., and T. Bache (1988). Locating events with a sparse network of regional arrays, *Bull. Seism. Soc. Am.* **78**, 780-798.

Douglas, A. (1967). Joint epicenter determination, *Nature*, **215**, 47-48.

Fisk, M. D. (2002). Accurate locations of nuclear explosions at the Lop Nor test site using alignment of seismograms and IKONOS satellite imagery, *Bull. Seismol. Soc. Am.* **92**, 2911-2925.

Hearn, T.M. and J.F. Ni (2001). Tomography and location problems in China using regional travel-time data, in *Proceedings, 23rd Seismic Research Review*, Jackson Hole, Wyoming, Oct. 2-5.

Jih, R. (1998). Location calibration efforts in China, in *Proceedings, 20th Seismic Research Symposium*, Santa Fe, New Mexico, Sept. 21-23.

Jordan, H.T. and K.A. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seism. Soc. Am.*, **71**, 1105-1130.

Kennett, B.L.N. and E.R. Engdahl (1991). Travel times for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429-465.

Kissling, W., W.L. Ellsworth, D. Eberhard-Phillips, and U. Kradolfer, Initial reference models in local earthquake tomography, *J. Geophys. Res.*, 99, 19635-19646, 1994.

Murphy, J.R., W.L. Rodi, M. Johnson, J.D. Sultanov, T.J. Bennett, M.N. Töksoz, C.E. Vincent, V. Ovtchinnikov, B.W. Barker, A.M. Rosca, and Y. Shchukin (2002). Seismic calibration of Group 1 International Monitoring System (IMS) stations in eastern Asia for improved event location, in *Proceedings, 24rd Seismic Research Review*, Ponte Vedra Beach, Florida, Sept. 17-19.

Ritzwoller, M.H., N.M. Shapiro, A.L. Levshin, E.A. Bergman, and E.R. Engdahl (2003), The ability of a global 3-D model to locate regional events, *J. Geophys. Res.*, 108, no. B7, 2353, ESE 9-1 - ESE 9-24.

Roecker, S.W., Seismicity and tectonics of the Pamir-Hindu Kush region of central Asia, *Ph.D. thesis*, MIT, Massachussetts, USA, 1977.

Sultanov, D. D., J.R. Murphy and Kh.D. Rubinstein (1999). A seismic source summary for Soviet peaceful nuclear explosions, *Bull. Seism. Soc. Am*, **89**, 640-647.

Waldhauser, F. and W.L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California, *Bull. Seism. Soc. Am.*, **90**, 1353-1368.

Waldhauser, F. (2001). HypoDD: A program to compute double-difference hypocenter locations, *U.S. Geol. Surv. open-file report*, **01-113**, Menlo Park, California.

Waldhauser, F., D. Schaff, P.G. Richards, and W.-Y. Kim (2004). Lop Nor revisited: Nuclear explosion locations, 1976-1996, from double-difference analysis of regional and teleseismic data, *Bull. Seism. Soc. Am.*, in press.

Yang, X., I. Bondár, K. McLaughlin, and R. North (2001). Source specific station corrections for regional phases at Fennoscandian stations, *Pure Appl. Geophys.*, **158**, 35-57.

Table 1: IMS (ABCE) stations considered in this study.

Station	Latitude	Longitude	Network
PS12 (HLR)	49.27	119.74	Primary
PS13 (LZH)	36.09	103.84	Primary
AS20 (BJI)	40.02	116.17	Auxiliary
AS21 (KMI)	25.15	102.75	Auxiliary
AS22 (SSE)	31.10	121.19	Auxiliary
AS23 (XAN)	34.04	108.92	Auxiliary

Table 2: Station distributions.

	Neijiang	Datong	Batang	Tangmai	Menyuan	Jinggu
Number of stations	36	64	67	58	17	13
Distance to nearest station (km)	120.5	210.4	125.5	274.9	86.3	191.4
Max. azimuth gap (°)	40.2	70.7	42.9	48.3	71.1	135.5

Table 3: Velocity models; Depth to top of layer in km; P-velocity, V*p*, in km/s.

Sichuan/Yunnan[1]		Tibet[2]		Gansu-Qinghai[2]		NE China, A[2]		IASP91	
Depth	V*p*	Depth	V*p*	Depth	V*p*	Depth	V*p*	Depth	V*p*
00.0	5.00	00.0	5.55	00.0	6.10	00.0	5.95	00.0	5.80
07.5	5.48	15.8	6.25	22.0	6.47	17.0	6.50	20.0	6.50
16.0	5.93	69.3	7.97	51.5	8.17	33.0	7.80	35.0	8.04
20.0	6.43								
30.0	6.60								
50.0	8.30								

[1] Z. Yang, personal communication.

[2] Jih (1998). Additional thinner layering is introduced to avoid sharp velocity contrasts.

Table 4: Reference Events at the GT5 and GT10(*) level

Date	Time	Lat	Lon	Depth	Mag	Location
1993/08/20	06:37:54.64	29.440	105.485	18.4	4.5	Neijiang
1994/04/14	17:57:06.24	29.441	105.511	14.7	4.4	Neijiang
1995/12/26	03:31:08.15	29.456	105.526	12.9	4.2	Neijiang
1996/01/16	00:08:15.86	29.408	105.467	06.7	4.2	Neijiang
1997/02/24	18:43:04.24	29.431	105.532	14.0	4.5	Neijiang
1997/08/13	08:13:29.83	29.473	105.587	11.2	4.8	Neijiang
1998/02/18	06:20:57.43	29.435	105.469	20.0	4.0	Neijiang
1999/08/17	10:41:05.03	29.425	105.541	07.9	4.8	Neijiang
1989/10/18	15:15:25.37	39.566	113.486	15.9	-	Datong
1989/10/18	17:01:33.24	39.568	113.463	13.2	-	Datong
1989/10/18	18:20:45.88	39.553	113.438	12.6	-	Datong
1989/10/18	19:37:49.36	39.592	113.486	13.5	-	Datong
1989/10/19	10:29:02.53	39.594	113.478	09.2	-	Datong
1989/10/19	12:32:15.66	39.544	113.474	17.2	-	Datong
1989/10/19	13:59:58.90	39.592	113.494	11.5	-	Datong
1989/10/19	17:56:47.80	39.579	113.505	09.9	-	Datong
1989/10/19	23:54:31.48	39.601	113.528	12.1	-	Datong
1989/10/20	11:41:41.62	39.581	113.542	13.0	-	Datong
1989/10/23	13:19:33.14	39.567	113.510	12.8	-	Datong
1989/10/23	17:07:54.28	39.557	113.488	11.8	-	Datong
1989/10/29	02:22:42.85	39.548	113.482	08.3	-	Datong
1989/12/08	13:05:14.00	39.549	113.484	12.8	-	Datong
1989/12/08	23:04:50.84	39.543	113.477	16.8	-	Datong
1989/12/31	08:24:48.13	39.559	113.499	12.4	-	Datong
1989/04/16	18:25:50.26	29.977	99.319	14.4	5.1	Batang
1989/04/25	02:13:22.08	29.993	99.320	17.9	6.2	Batang
1989/04/30	23:05:26.82	29.964	99.353	17.2	5.1	Batang
1989/05/03	05:53:00.94	30.020	99.362	14.4	6.1	Batang
1989/05/03	15:41:30.94	29.966	99.365	14.3	5.8	Batang
1989/05/03	17:28:21.44	30.047	99.354	12.5	5.3	Batang
1989/05/04	05:30:46.27	29.963	99.364	10.1	5.1	Batang
1985/07/18	17:40:13.52	30.338	94.799	28.1	4.9	Tangmai*
1985/07/19	02:38:08.63	30.323	94.810	16.2	4.7	Tangmai*
1985/07/20	18:31:45.12	30.319	94.818	21.6	4.6	Tangmai*
1986/10/10	08:59:19.47	30.325	94.773	18.5	4.8	Tangmai*
1986/10/12	16:29:11.50	30.343	94.810	25.7	4.7	Tangmai*

Table 4: (Continued)

1987/09/17	01:34:47.20	30.330	94.788	15.2	4.9	Tangmai*
1987/09/19	18:59:38.80	30.323	94.806	29.3	4.7	Tangmai*
1991/07/18	13:25:00.18	30.320	94.802	25.6	5.0	Tangmai*
1991/07/20	18:52:24.44	30.309	94.802	28.6	4.5	Tangmai*
1991/07/20	19:02:31.24	30.321	94.808	18.6	4.8	Tangmai*
1991/07/23	16:51:53.69	30.323	94.801	18.5	4.7	Tangmai*
1991/07/24	06:06:45.16	30.306	94.794	27.4	4.8	Tangmai*
1991/07/25	01:52:44.28	30.315	94.782	12.0	4.8	Tangmai*
1991/07/28	23:58:20.64	30.331	94.760	29.7	4.9	Tangmai*
1991/07/29	03:20:16.09	30.305	94.816	28.1	4.6	Tangmai*
1993/09/06	20:57:22.32	30.312	94.816	16.1	4.7	Tangmai*
1986/08/26	09:43:00.63	37.796	101.669	13.2	6.2	Menyuan
1986/08/26	10:30:00.03	37.753	101.688	15.7	5.4	Menyuan
1986/08/26	13:11:24.61	37.753	101.673	10.6	5.0	Menyuan
1986/08/27	13:15:23.03	37.704	101.678	10.7	4.3	Menyuan
1987/06/28	01:16:36.88	37.731	101.672	14.0	4.9	Menyuan
1993/05/30	10:01:10.33	23.715	100.501	11.5	-	Jinggu*
1993/05/30	21:49:02.14	23.746	100.483	13.0	-	Jinggu*
1993/06/04	01:04:00.94	23.707	100.500	11.4	4.5	Jinggu*
1993/06/10	20:38:28.68	23.691	100.490	12.2	4.6	Jinggu*
1993/10/25	08:32:46.55	23.582	100.523	5.9	-	Jinggu*
1994/03/18	16:16:47.64	23.606	100.521	5.7	-	Jinggu*
1994/11/07	22:40:50.12	23.720	100.464	12.9	4.0	Jinggu*

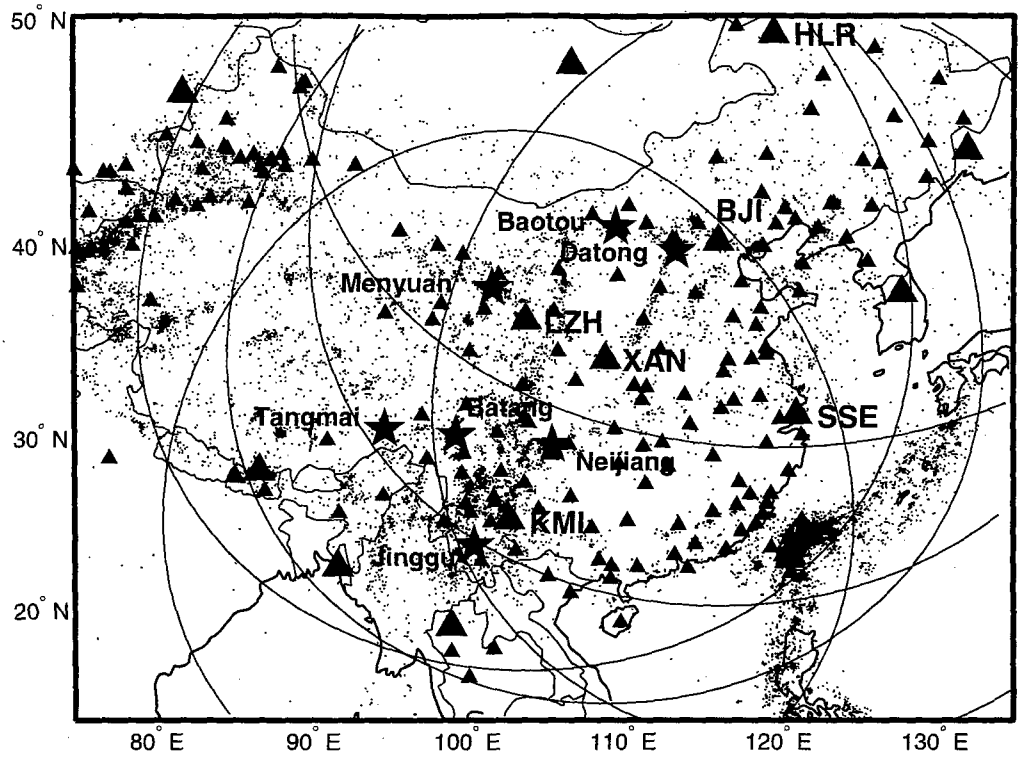

Figure 1 Events (dots) and Chinese stations (small triangles) listed in the Annual Bulletin of Chinese Earthquakes (ABCE). IMS stations (large triangles) are indicated and the ones discussed in this study labeled. Stars indicate locations at which reference events are obtained. Circles denote areas out to 20° from each of the labeled IMS stations.

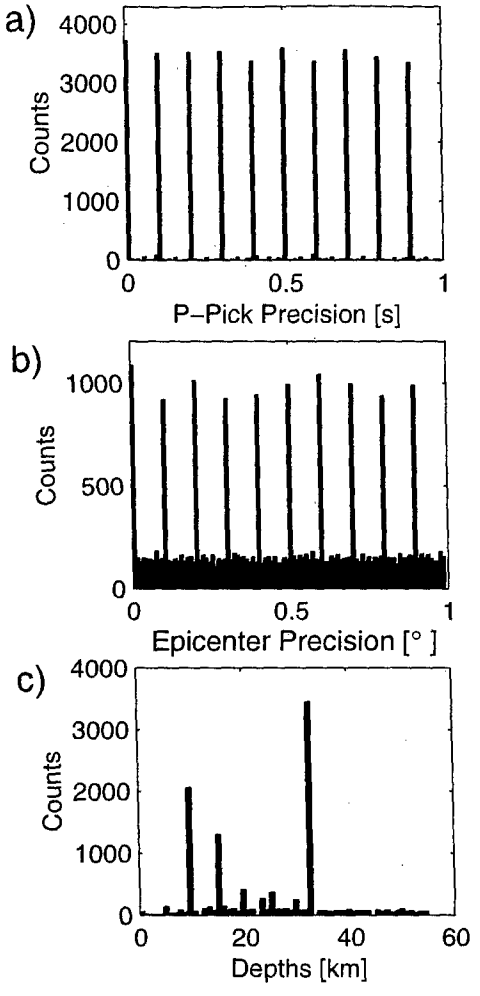

Figure 2 Reporting precision of (a) ABCE *P*-phase picks and (b) epicenter locations. In a) the distribution is shown for fractions of seconds, in b) fractions of degrees. (c) Distribution of event depths.

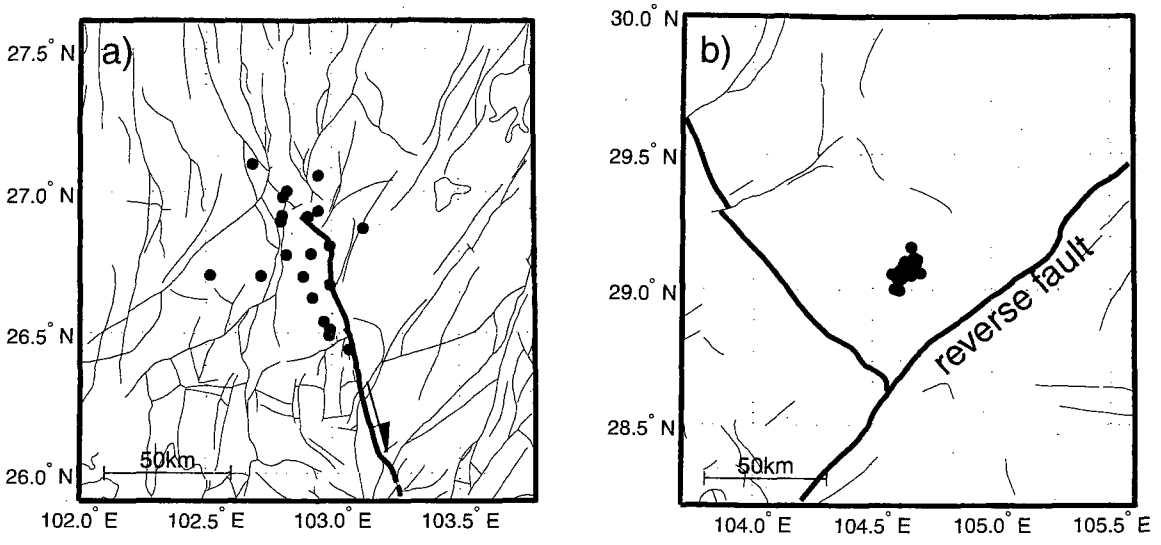

Figure 3 Two examples where no unique association of the relocated seismicity with an active fault is possible. a) 21 events, that occurred between 1987-1998, do not appear to occur on a single deep fault, but rather on a set of adjacent smaller faults. b) 38 events, between 1985-1996, are located more than 40 km from the nearest mapped fault.

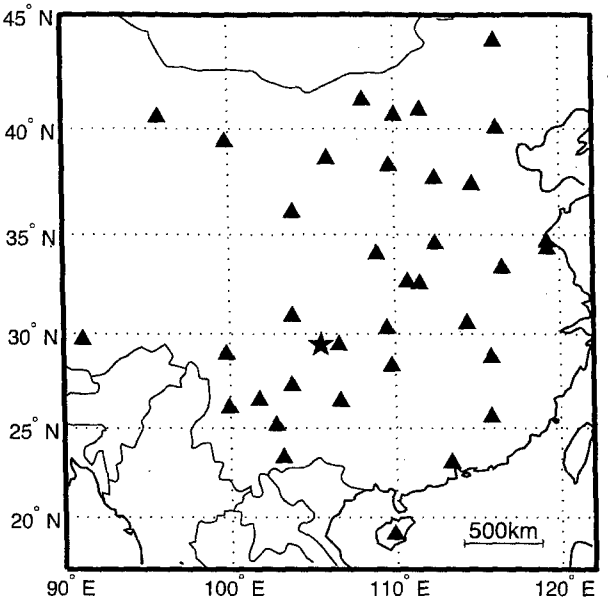

Figure 4 Stations (triangles) that recorded the 61 events near Neijiang, Sichuan Province. Star indicates the location of the cluster.

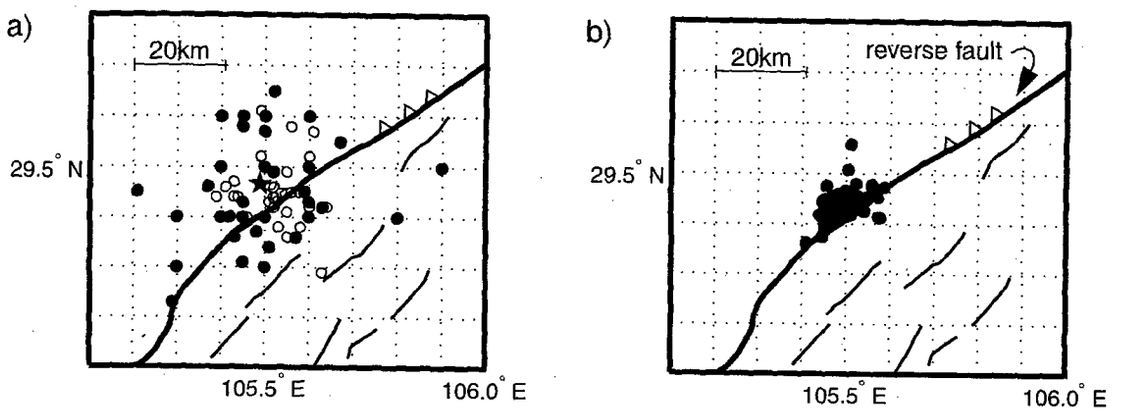

Figure 5 Map view of (a) ABCE locations and (b) double-difference locations. Solid circles in (a) indicate locations from the electronic ABCE, open circles those from the printed ABCE. Solid and open stars represent the centroid of the electronic and printed ABCE locations, respectively. Lines indicate mapped surface traces of deep reverse faults.

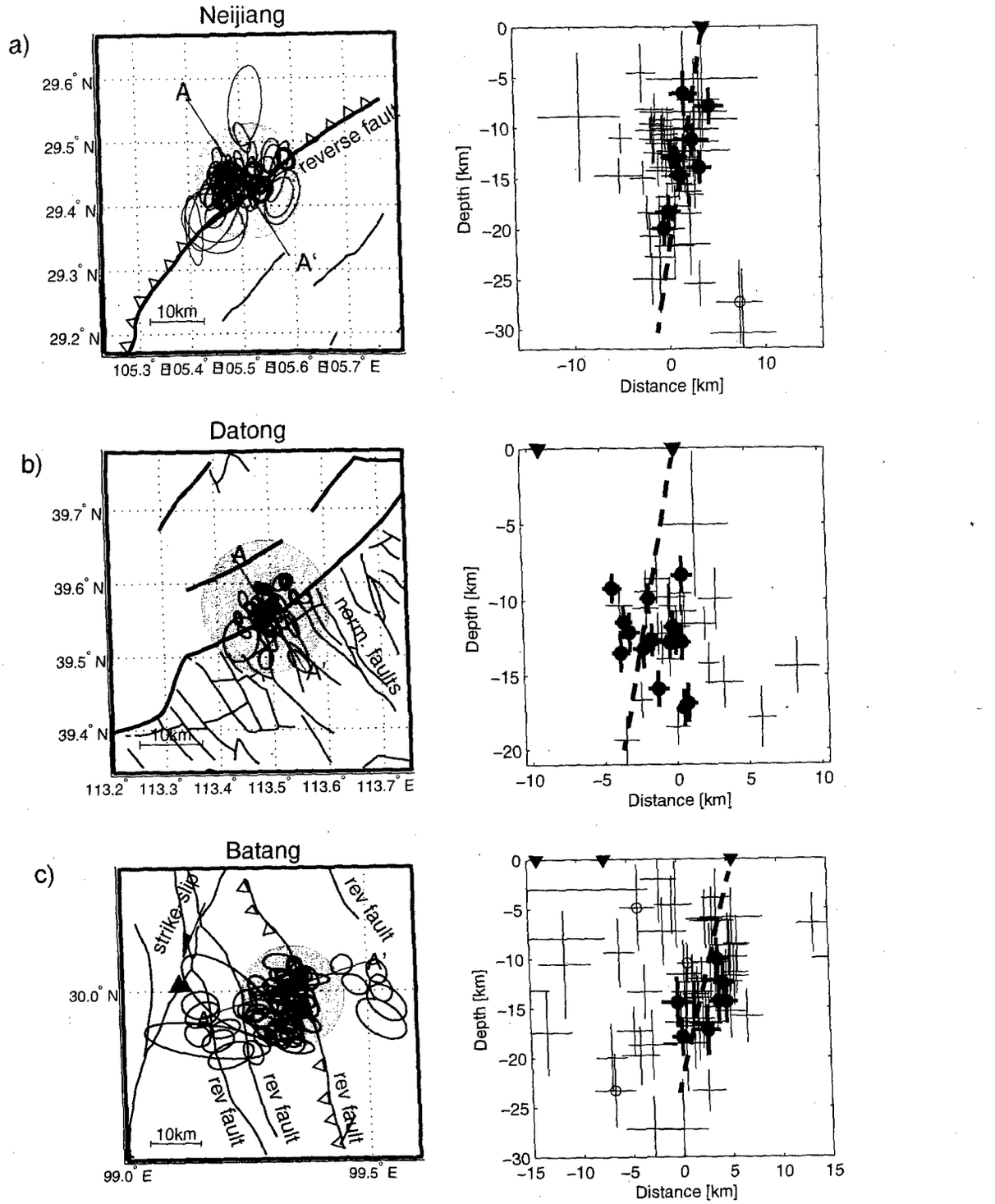

Figure 6 Relocated events in map view (left panel) and fault perpendicular cross section (right panel) along A-A' for events near (a) Neijiang, (b) Datong, (c) Batang, (d) Tangmai, (e) Menyuan, (f) Jinggu. Ellipses (in map view) and crosses (in cross sections) indicate bootstrap errors at the 90% confidence level. Reference events are indicated by thick ellipses or crosses. Lines in map view indicate mapped faults at the surface, thick lines those that are described as deep faults in the USGS fault data base. Faults are labeled as indicated in the fault data base. Triangles denote the projection of the main fault onto the cross section. Dashed line in cross section indicates assumed fault dip based on seismicity. Open circles denote events with $M \geq 3.5$, except in the Batang cluster ($M \geq 5$), and the Tangmai cluster ($M \geq 4.5$). No magnitudes are available for events in the Datong cluster. Gray filled circles in map view are centered on the cluster centroid and indicate the estimated average absolute location uncertainty associated with the ABCE locations. rev. = reverse; norm. = normal faulting.

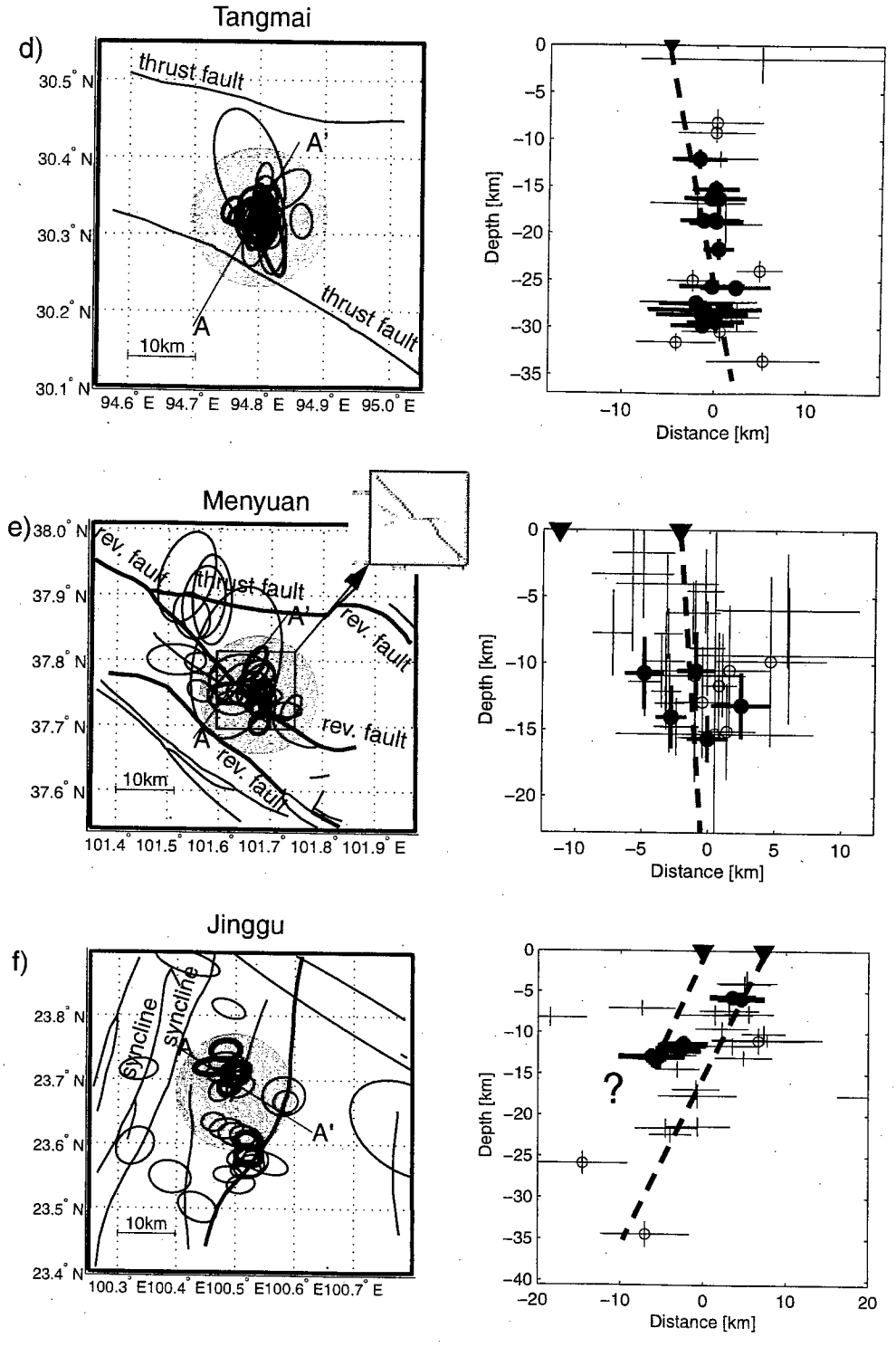

Figure 6 (Continued)

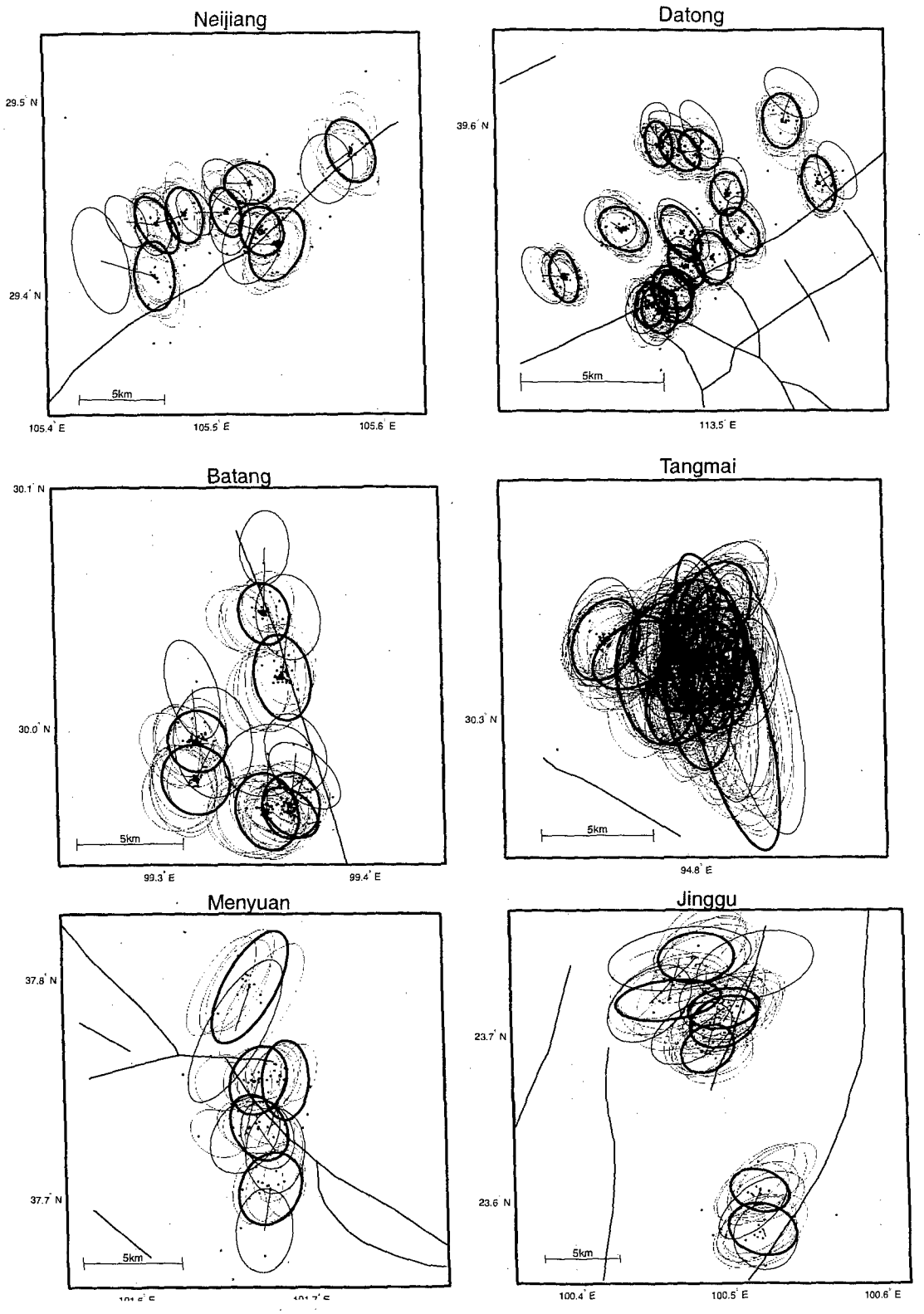

Figure 7 Results from three different tests performed to ensure robustness of the seismicity structure within each cluster. Thick ellipses represent final locations of reference events, thin ellipses show relocation results from using the IASP91 velocity model, gray ellipses are from relocating the events with one event removed at a time, dots from relocating the events with one station removed at a time. Lines indicate fault traces, except for lines that connect the final reference events with locations derived with IASP91 model. See text for explanation.

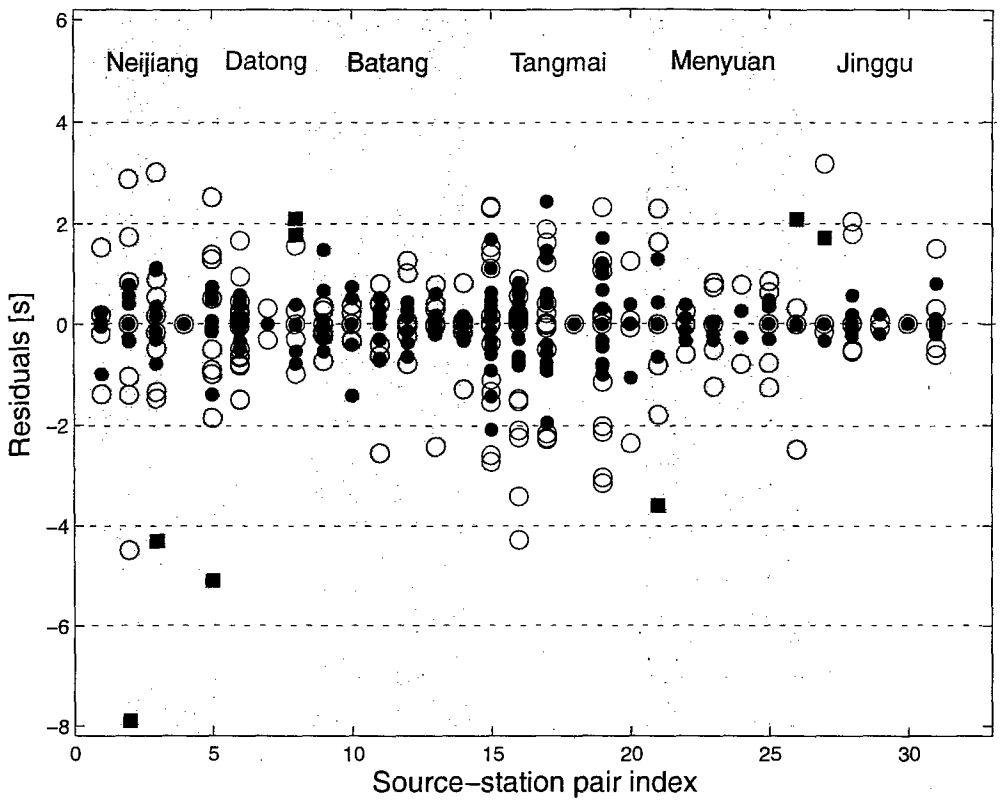

Figure 8 P-wave travel time residuals relative to the median travel time shown for each cluster-station pair (see Figure 9). Solid circles indicate travel time residuals from reference events, open circles those from the ABCE corresponding locations. Squares indicate phase picks that are considered outliers. Standard deviation for reference event residuals, after removing the outliers, is 0.61 sec, for ABCE residuals 1.28 sec.

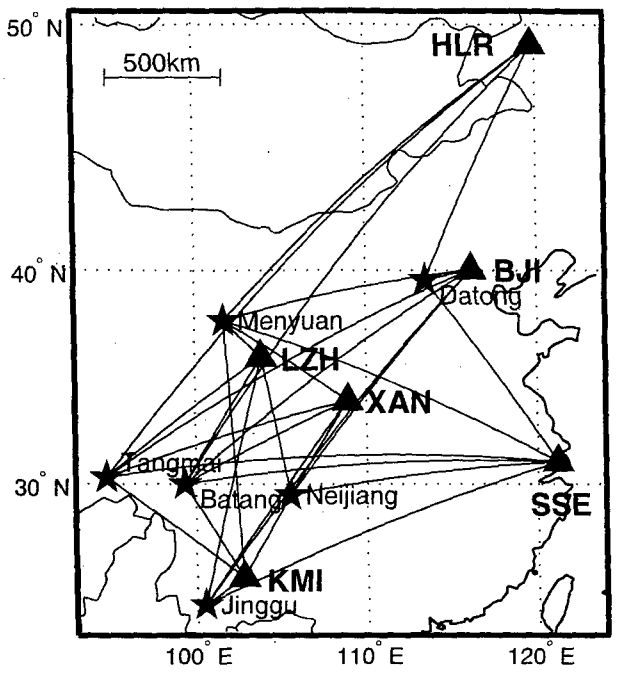

Figure 9 Great circle paths of P-waves generated by clustered reference events determined in this study (stars) and recorded at six IMS stations in China (triangles), providing useful calibration data. See Table 1 for station coordinates, and Table 4 for reference event locations.

SECTION TWO

Double-difference Relocation of Earthquakes in central-western China

The double-difference earthquake location algorithm (Waldhauser and Ellsworth, 2000) was applied to the relocation of 10,057 earthquakes that occurred in central western China (21°N to 36°N, 98°E to 111°E) during the period from 1992 to 1999. In total, 79,706 readings for *P*-waves and 72,169 readings for *S*-waves were used in the relocation. The relocated seismicity (6,496 earthquakes) images fault structures that are in close correlation with the tectonic structure of major fault systems expressed at the surface. This is the first study of a region of major seismicity in China for which earthquake locations are concordant with mapped active faults.

The original data set was assembled by seismologists in the China Earthquake Administration, who had access to phase picks generated by seismographic networks in Sichuan, Yunnan, Shaanxi, and Guangxi Provinces; and to phase picks from the China National Seismographic Network. From these separate networks, 15,092 earthquakes had been located, as shown in Figure 1. After merging the phase picks, they were associated with 10,057 separate events and these were the input to the double-difference algorithm. The relocations are shown in Figure 2. A full account of the comparison between relocated events and active fault traces is given by Yang et al. (2005).

References

Waldhauser, F., and W.L. Ellsworth, A double difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California, *Bulletin of the Seismological Society of America*, **90**, 1353--1368, 2000.

Yang, Z.X., F. Waldhauser, Y.T. Chen, and P.G. Richards, Double-difference relocation of earthquakes in central-western China, 1992--1999, *Journal of Seismology*, **9**, 241--264, 2005.

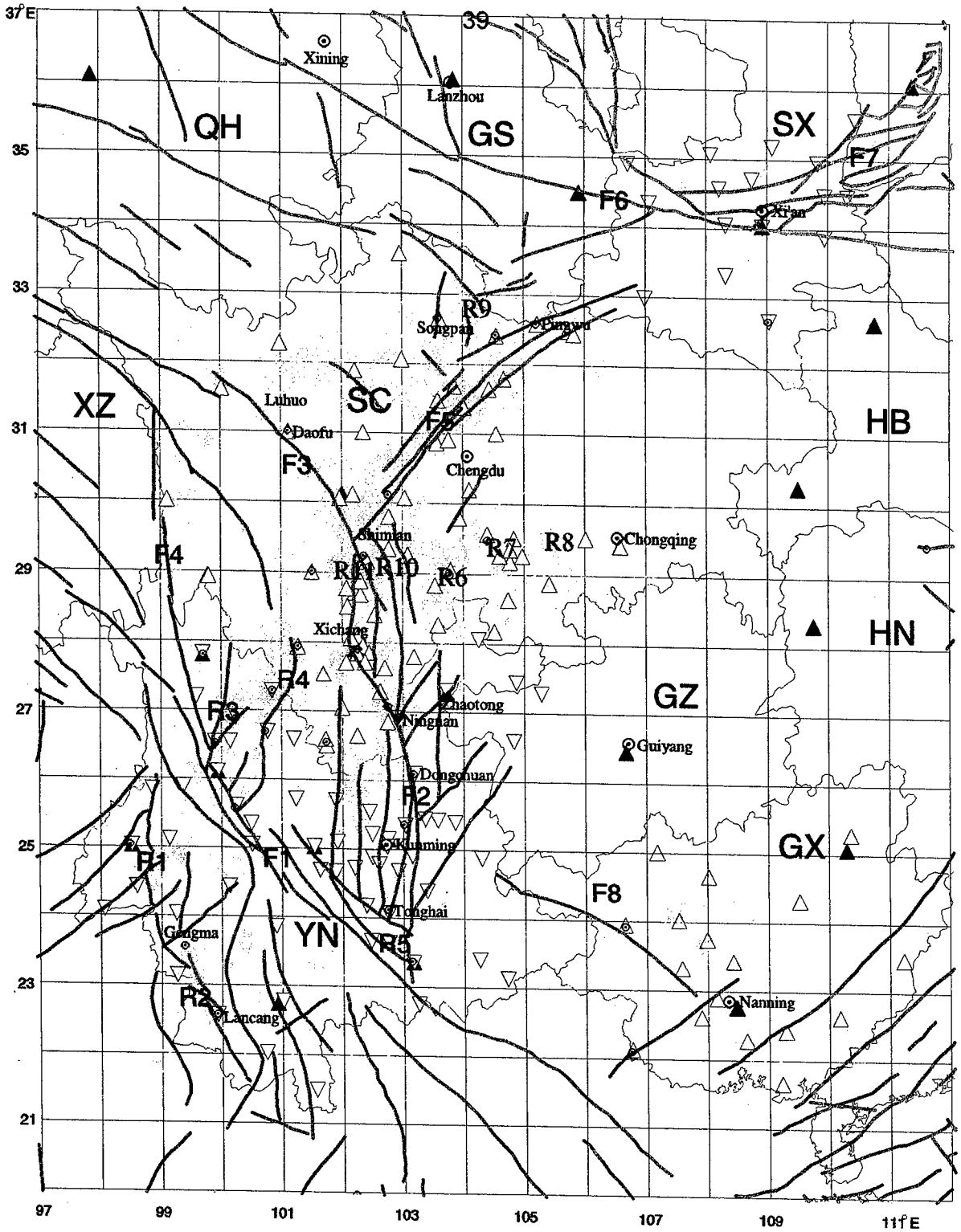

Figure 1. Epicentral distribution of all the 15,092 routinely located earthquakes (gray circles) from 1992 to 1999. Thick lines represent surface traces of active faults. Thin lines delineate provincial boundaries. Triangles show station locations of provincial networks (grey and open), and the national network (black).

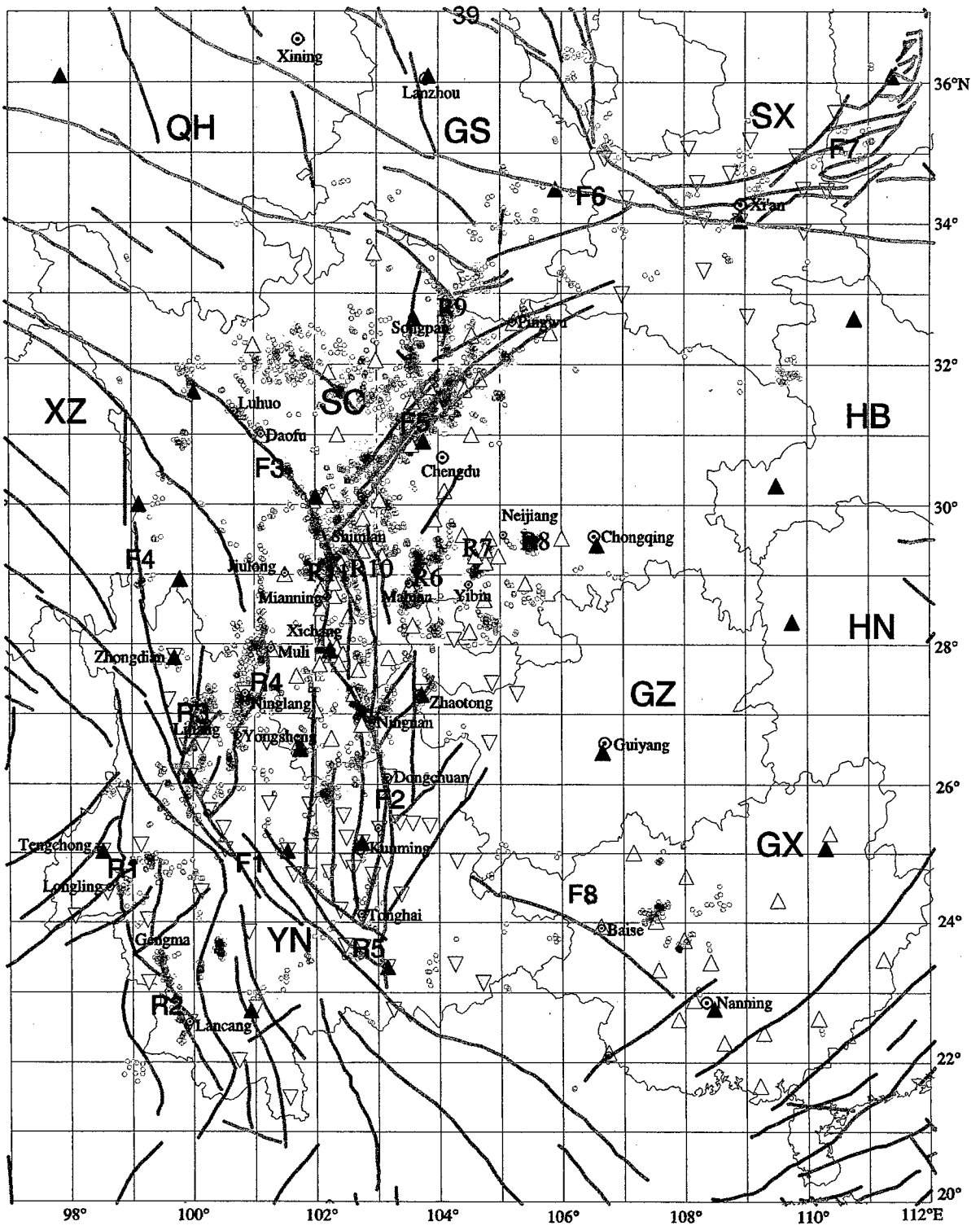

Figure 2. Epicentral distribution of all 6,496 relocated earthquakes in central-western China using the double difference algorithm. Gray circles represent the same events before relocation. Other symbols as in Figure 1. There is much better agreement between relocated events and fault traces, than between original locations and fault traces.

SECTION THREE

The Applicability of Modern Methods of Earthquake Location

PAUL G. RICHARDS[1, 2], FELIX WALDHAUSER[1], DAVID SCHAFF[1], and WON-YOUNG KIM[1]

Abstract—We compare traditional methods of seismic event location, based on phase pick data and analysis of events one-at-a-time, with a modern method based on cross-correlation measurements and joint analysis of numerous events. In application to four different regions representing different types of seismicity and monitored with networks of different station density, we present preliminary results indicating what fraction of seismic events may be amenable to analysis with modern methods.

The latter can supply locations ten to a hundred times more precise than traditional methods. Since good locations of seismic sources are needed as the starting point for so many user communities, and potentially can be provided due to current improvements in easily-accessible computational capability, we advocate wide-scale application of modern methods in the routine production of bulletins of seismicity. This effort requires access to waveform archives from well-calibrated stations that have long operated at the same location.

Key words: Earthquake location, waveform cross-correlation, seismicity studies, California earthquakes, Charlevoix earthquakes, China earthquakes, New Madrid earthquakes

---

[1]Lamont-Doherty Earth Observatory, 61 Route 9W, Palisades, New York 10964, USA

[2]also Department of Earth and Environmental Sciences, Columbia University; richards@LDEO.columbia.edu

Introduction

Seismic events are usually still located one-at-a-time by measuring the arrival times of different seismic signals (phase picks) and then interpreting these observations in terms of the travel times predicted for a standard depth-dependent Earth model. In this traditional approach, the differences between observed and calculated arrival times (based on a trial origin time and location) are reduced by a process of iteration (for each event separately) to a value deemed acceptable.

Many different studies of specific regions and particular datasets have demonstrated that by use of whole waveforms and locating groups of events all together, location estimates can be very significantly improved over the results obtained by the traditional approach. In this paper we loosely refer to analysis of waveforms, and joint location of many events, as "modern methods"—in contrast to "traditional methods" based on phase picks and location of events one-at-a-time. We examine practical aspects of the earthquake location problem for four different regions, and assess the merits of a modern method in which waveform cross-correlation is applied to large data sets. We find the fraction of events in each region that are amenable to this type of relocation, and comment on the degree of location improvement.

In regions for which an extensive archive of waveforms can be obtained from stations with a long history of operation at the same site with the same instruments, modern methods of event location (i.e. those that combine waveform cross-correlation with a multi-event location algorithm) may now be considered seriously for application to the routine publication of seismicity bulletins.

Sections that follow give background on traditional and modern methods, describe in general our method of analysis for the different regions studied, and present results for the different regions. We conclude with discussion and comment on traditional versus modern methods of event location.

This is not a review paper that compares different methods of cross correlation and event relocation to come up with a verdict about which combination works best in a particular area.

The practical experience of network operators will typically be an important guide in determining that combination for different areas. Nor is this an interpretational paper, in which results from each of the four regions are discussed in terms of their implications on the tectonics or earthquake physics. Rather, it is our purpose to present results on relocation using a modern method using waveform cross-correlation to the extent permitted by available data and relocating multiple events at the same time; and to comment upon the degree of location improvement, in so far as possible using a common framework of analysis but applied to a broad variety of tectonic regions at different scales, and using seismographic monitoring networks of varying station density.

Background

Seismic data derived from earthquakes and explosions are used in scientific studies of Earth's internal structure, tectonics, and the physics of earthquake processes; in engineering studies of earthquake hazard and efforts in mitigation and emergency management; and in monitoring of nuclear explosions, either as a military activity to evaluate the weapons-development program of a potential adversary, or as an arms-control initiative such as monitoring compliance with the Comprehensive Nuclear-Test-Ban Treaty.

In practice the great majority of those who work with seismic data do not use seismograms directly. Instead they typically use data products derived from seismograms. The most important of these products are bulletins of seismicity, which consist of catalogs of earthquake and explosion locations, measures of event size (such as magnitudes, and scalar and tensor moments), and associated data such as the arrival times of seismic waves at particular stations.

Since the late 1970s there have been enormous improvements in the quality and quantity of seismograms, associated with the deployment of broadband feedback sensors and techniques

of digital recording that can faithfully capture signals with high dynamic range across wide bands of frequency; and there is ongoing revolutionary improvement in access to seismogram data, handicapped only by political barriers as reliable satellite communications and the Internet spread even to remote locations and become less expensive. It may therefore seem surprising that routinely published global and regional bulletins of seismicity have not yet seen a corresponding radical improvement that would greatly benefit any of the principal user communities, even though demonstrably successful modern methods of event location and source characterization have been developed and applied in numerous special studies. As discussed further, below, the main reason modern methods have yet to be widely applied is that it is necessary to build up major archives of well-calibrated and easily accessible waveforms from fixed stations operated over many years.

When teleseismic arrival times are used for event location, the resulting location estimates using traditional procedures are typically in error by several km for events detected at hundreds of stations, and by a few tens of km for events detected at tens of stations. Such errors in traditional event locations in standard bulletins are not always appreciated by users, but can be shown for example by relocation studies such as that of ENGDAHL et al. (1998), and by comparison of standard location estimates with the hypocenters of events whose ground truth location is known from non-seismic methods (BONDÁR et al., 2004). It can also be demonstrated by study (SCHAFF and RICHARDS, 2004a) of waveform doublets that must be within about 1 km of each other, but that standard bulletins report as having been tens of km apart when the events are located in the usual way (i.e., one-at-a-time from phase picks).

When regional arrival times are used, there is practical experience in the western U.S. to indicate that location uncertainties (as given in conventional bulletins) are at the level of one or two km in areas with high station density (station separation on the order of 10–20 km); and are at the level of about five km in areas with fewer stations (station spacing, approximately 50 km) (personal communication, HAUKSSON, 2005). In broad areas for which events are located with regional signals on the basis of a very sparse set of stations, mislocations can routinely reach

several tens of km (RICHARDS et al., 2003).

There are two independent reasons why events can be significantly mislocated, even in situations where station coverage is not a problem. First, there is the difficulty of picking arrival times accurately. When signal-to-noise ratios are good, the error in traditional methods of measuring the arrival time of seismic waves is usually less than one second (and can be less than 0.1 second in favorable conditions). But signals (*Pn*, for example) are often weak and/or emergent, and later arrivals have to be picked in time windows that include the coda of earlier arrivals. Thus regional *S*-wave phases may be picked with errors that in some cases reach up to a few tens of seconds. Second and often more important, there is the problem of errors in the travel-time model used to interpret measured arrival times. In many cases it is model error, not pick error, which dominates the overall error in absolute location. In such cases the overall goal of improving locations based on seismic arrival-times can be achieved only by reducing the effect of model errors. Model errors can also lead to poor-quality relative locations of events that in fact lie near to each other, if the events are located one-at-a-time using data derived from different station sets.

At depths greater than about 200 km, the Earth's laterally-averaged velocity structure is known quite accurately (i.e. to well within 1% at most depths, the biggest difference from actual velocities being in regions of subducting tectonic plates). The main difficulty is at shallower depths, within the crust and uppermost mantle, where the actual velocity of seismic waves may differ in unknown ways, perhaps by as much as 10%, from the velocity that is often assumed (such as the velocity given by *ak135* or some other standard Earth model). Thus the problem of substantial differences between actual and standard travel-time models is most significant for regional waves. If a seismic event is about 500 km from a station that detects a regional arrival, and if the arrival is misinterpreted with a velocity that is wrong by 5%, then the event will be estimated from that observation alone as having originated at a distance that is incorrect by about 25 km. Non-horizontal interfaces within the Earth, and phase misidentification (for example between *Pg* and *Pn*) compound these problems.

Experience with the traditional location method originated in the era of analog recording, with sensors of limited dynamic range and bandwidth. It is therefore remarkable that even after decades of experience with digital seismic data beginning in the 1970s, event location is usually still based upon analysis of one-event-at-a-time, and traditional picking of the arrival times for specific seismic phases. In practice this means that only a very small fraction of the available information contained in a modern seismogram is being used to locate events.

In this paper we use a recently-developed highly-efficient waveform cross-correlation algorithm for locally-observed phases (SCHAFF et al., 2004) to measure accurate arrival-time differences for neighboring earthquakes observed at common stations. The waveform cross-correlation method has been applied on the local scale with considerable success. It uses much more of the information contained in seismograms, and can significantly reduce the error in relative arrival times derived from phase picks. We analyze cross-correlation measurements on local and regional signals, combined with ordinary phase picks from earthquake bulletins, using the double-difference technique (WALDHAUSER and ELLSWORTH, 2000; WALDHAUSER, 2001) to obtain precise relative locations for many earthquakes all at the same time.

Several other algorithms have been developed to reduce the effect of errors in the travel time model by locating many events at the same time. These include joint hypocenter determination (JHD) (DOUGLAS, 1967; and FROHLICH, 1979, among others), hypocentroidal decomposition (HDC) (JORDAN and SVERDRUP, 1981), or the use of source specific station terms (SSST) (RICHARDS-DINGER and SHEARER, 2000). All of these methods attempt to correct for correlated effects, due to three-dimensional Earth structure, in the arrival times generated by a cluster of events. The corrections are typically accomplished via use of residual-based station corrections, either in the form of a 1D (static) correction (JHD, HDC), or by using source specific station terms (SSST). The former reduces model errors due to unusual structure beneath a station, while the latter reduces model errors due to structure somewhere between the region of seismicity for which a correction factor is determined, and a particular station. With the double-difference method (DD), a network of events is built in which each event is linked to its nearest

neighbors through travel-time differences observed at common stations, which are directly inverted for event separation. The method thus differs from earlier methods of joint hypocentral determination in that no station corrections are necessary, because unmodeled velocity structure is directly removed for each event pair individually. Thus double-difference results are best if events are densely distributed (a few km event separations when using local stations separated by several tens of km). For studies based on regional signals (distances typically in the range from a few hundred km up to about a thousand km), events can be more sparse although event pairs need to have approximately the same take-off direction for corresponding wave types recorded at any common station. The web of DD-relocated events can be expected to give precise relative locations for neighboring events, particularly when pick error is avoided by use of cross-correlation. It is not our purpose here to compare these different algorithms, which we would expect to give similar results for clusters of events. Waveform cross-correlation methods together with multi-event location algorithms have been applied by many previous investigators (for example POUPINET et al., 1984; ISRAELSON, 1990; HARRIS, 1991; GOT et al., 1994; WALDHAUSER et al., 1999 and 2004ab; RUBIN et al., 1999; WITHERS et al., 1999; THURBER et al., 2001; SCHAFF et al., 2002; ROWE et al., 2002; XIE et al., 1997, and XIE, 2001; SHEARER, 1997, HAUKSSON and SHEARER, 2005, and SHEARER et al., 2005). Most of these studies focused on local earthquakes, though a few used regional and teleseismic waveforms to locate small clusters of events. These papers have convincingly demonstrated major improvements in the precision of relative (and in some cases absolute) event locations. In the present paper we are interested in evaluating a modern method (i.e., waveform cross-correlation to the extent permitted by available data, combined with multiple event location) on different scales, including some examples at quite large scales; and in commenting generally on the advantages for the modern approach that accrue at larger scales.

Method of Analysis for the Different Regions

For each of the four regions of seismicity to which we have applied modern methods of event location, we have taken the three steps of (a) data collection, (b) waveform analysis, and (c) event relocation using the double difference (DD) algorithm.

The four regions of seismicity for which we have obtained relocations and on which we comment in this paper are:

(1) A broad-area study of the seismicity of China and surrounding regions from 1985 to 2000, described in more detail by SCHAFF and RICHARDS (2004a). This study built upon a relatively small study of foreshocks, main shock, and aftershocks, over a period of several months for the Xiuyan region of Liaoning Province, China, described in more detail by SCHAFF and RICHARDS (2004b). The small study used regional seismograms in the distance range 600 to 1500 km archived by the IRIS Consortium's Data Management Center (IRIS DMC), and showed that complex *Lg* waveforms correlated well for a significant fraction of events for time windows of several hundred seconds duration. The larger study of China was based on 115 stations at distances up to 20° from about 14,000 events, with measurements made on about 130,000 seismograms. Cross-correlation was based on window lengths from 60 to 400 s beginning prior to the first *P*-arrival, ending 40 s after *Lg*, and signals passed in the band from 0.5 to 5 hz. 1.2 million cross-correlations were made, for signals associated with event pairs that (according to the Annual Bulletin of Chinese Earthquakes) were not more than 150 km apart.

(2) A study of the New Madrid region of the central United States, for two different eras of station deployment. Data consisted of phase picks as well as waveforms (used for cross-correlation measurement), associated with seismicity bulletins of events located by traditional methods applied by local network operators. These were the Center for Earthquake Research Information of the University of Memphis, which deployed a PANDA network (42 stations, 918 events, 17,598 phases) from October 1989 to August 1992, and a different network (85 stations, 614 events, 16,461 phase picks) from January 2000 to October 2003. Window lengths of 1 s

were used for cross-correlation measurements.

(3) A study of the Charlevoix region of Eastern Canada from January 1988 to December 2003 involving 2797 events recorded at 46 stations operated by the Geological Survey of Canada with 33,423 phase picks. Window lengths of 1 s were used for cross-correlation measurements. In practice, almost all of the highly cross-correlated seismograms derived from only 7 stations, closest to the Charlevoix region.

(4) A study of the region monitored by the Northern California Seismic Network (NCSN). This was by far the largest study, with about 225,000 events recorded by 900 stations from 1984 to 2003. Window lengths of only 1 and 2 seconds were used, and the NCSN seismicity bulletin (prepared using traditional methods) was used to select these time windows for all events whose inter-event distance was up to 5 km. Further details are given by SCHAFF and WALDHAUSER (2005).

The associated waveforms for sets of previously defined events were obtained for study (1) from the IRIS DMC; and for study (4) from the Northern California Earthquake Data Center (including phase picks in this case). Though large amounts of data were requested, these two acquisitions were straightforward in that both these data centers had archives previously prepared with the expectation that outside scientists would be using them. But for studies (2) and (3) the basic datasets of associated waveforms and phase picks entailed considerable effort to assemble (with the willing help of network operators). It was necessary to account for different formats, different station sets, and different practices of bulletin preparation over the periods of time during which the seismicity in these regions had been documented.

For some purposes, such as scientific study of earthquake interactions in a fault zone or seismic sources associated with magma conduits in a volcano, relative locations can be sufficient. But of course for seismic monitoring of nuclear explosions, or identification of earthquakes with particular fault structures, there is the need to move beyond relative locations and to estimate absolute hypocentral coordinates. In this paper we focus on improvements in precision, i.e. on relative locations, noting that absolute locations may be achievable at a later

step in several different ways—for example by use of ground truth information for a subset of events, and/or by use of a good 3D travel-time model.

Preliminary Results for Four Different Regions

(1) China and surrounding regions of East Asia.

This project began with a small preliminary study of 28 earthquakes drawn from a sequence of 90 events during 1999 – 2000 in the Xiuyan region of Liaoning Province, and moved on to an analysis of about 14,000 events for all of China.

The small study surprisingly showed that in some cases the complex, highly scattered *Lg*-wave is similar at far-regional distances for clusters of events (SCHAFF and RICHARDS, 2004b). Figure 1 shows the high degree of similarity for more than 300 s of waveform, even at frequencies up to 5 hz. Cross correlations provided highly accurate differential travel-time measurements. Their error estimated from the internal consistency is about 7 ms. These travel time differences were then inverted by the double-difference technique to obtain epicenter estimates that have location precision on the order of 150 meters, as shown in Figure 2. The locations were computed with waveform data acquired by four or five regional stations 500 to 1000 km away. The relative epicenter estimates were not substantially affected by the paucity of stations or by large azimuthal gaps. For example, using four stations with an azimuthal gap of 133°, or using only three stations with an azimuthal gap of 220°, resulted in locations that differed by less than 150 m.

Regional event locations must often be based on a small number of phases and stations due to weak signal-to-noise ratios and sparse station coverage. This is especially true for monitoring work that seeks to locate smaller magnitude seismic events with a handful of regional stations. Two primary advantages of using *Lg* for detection and location are that it is commonly the largest amplitude regional wave (enabling detection of smaller events); and it propagates more slowly than *P*-waves or *Sn* (resulting in smaller uncertainty in distance, for a given

uncertainty in travel time). This preliminary study demanded a high standard for identification of similar events (cross-correlation ≥ 0.8 for a several hundred sec window of signal passed in the band from 0.5 to 5 hz).

In the larger study we found that about 9% of seismic events in and near China, from 1985 to 2000, were repeating events not more than about 1 kilometer from each other (SCHAFF and RICHARDS, 2004a). This conclusion was based on the stringent criterion for waveform similarity described in Figure 3 (see caption). We cross-correlated seismograms from about 14,000 earthquakes and explosions and measured relative arrival times to within about 0.01 seconds, enabling lateral location precision of about 100-300 meters. Recognition and measurement of repeating signals in archived data and the resulting improved locations enabled us to quantify the inaccuracy of current procedures for picking onset times and locating events. The fraction of cross-correlated events, though quite low (9%) in Figure 3 for the whole time period from 1985 to 2000, rose to a larger value (14%) in the later years, as shown in Figure 4. Table 1 shows too that the fraction increased significantly if the stringent criterion (CC ≥ 0.8) is relaxed. Thus, relaxation to CC ≥ 0.6 resulted in 23% of the events in the ABCE cross correlating successfully for the whole time period.

(2) New Madrid, Central United States.

This region of intraplate seismicity in the Central United States experiences about 200–250 locatable earthquakes each year. It has been monitored by various networks since the 1970s, and currently we have preliminary relocation results for one three-year period and one four-year period.

Thus, Figure 5 shows relocation of 783 events monitored by a 42-station network from 1989 to 1992. Although 707 of the events (90%) had five or more *P*-wave cross-correlations (CC ≥ 0.7), these were not necessarily to the same event. It is possible that these events are relocated with high precision, but this is not assured. We found that a subset of 616 events (65%) had five or more *P*-wave cross-correlations (CC ≥ 0.7) to a neighboring event and were thus

relocated with high precision provided the azimuthal distribution of stations was adequate; 695 events (85%) cross-correlated at four or more stations; and 735 (90%) at three or more stations. Figure 6 shows relocation of 594 events that occurred in the period January 2000 to October 2003 and were monitored by an 82-station network still operating in 2005. In this case, 499 events (84%) had five or more *P*-wave cross-correlations (CC $\geq$ 0.7), but these were not necessarily to the same event. 371 events (63%) had five or more *P*-wave cross-correlations (CC $\geq$ 0.7) to a neighboring event and were thus relocated with highest precision in our study. Events that do not cross correlate are still located more precisely than in the traditional bulletin, because of the use of the double-difference algorithm applied to phase-pick data for the whole set of events.

The general distribution of seismicity is very similar between Figures 5 and 6. Figure 7 shows the locations obtained by traditional methods for the same period and station set used in Figure 6. It is apparent especially in the cross sections and in the map view of events in the southwest sub-region, that the relocated events in Figure 6 more clearly identify lineations than do the traditionally-obtained locations of Figure 7.

(3) Charlevoix, Eastern Canada.

This too is a region of intraplate seismicity. The preliminary relocations shown in Figure 8, as well as the original bulletin locations for this region, indicate that active faulting is simpler and more clearly defined in the northeastern part of the Charlevoix seismic zone, compared to more complex features in the southwestern part. The dominant faults are two parallel southwest-northeast running structures that dip to the southeast. Of these, the more eastern fault has a dip of about 50° (cross-section 2–2′) and that to the northwest has a dip of about 75° (cross-section 3–3′). In the northeast the seismicity starts at about 7 or 8 km depth; in the southwest it starts at about 4 km depth. Seismicity extends down to about 30–35 km depth. Larger events occur in the northeastern sub-region (circled events have magnitude $\geq$ 4).

For the Charlevoix seismicity shown as 2272 relocated events in Figure 8, only 242 of

them (5%) cross-correlated (with CC ≥ 0.7) at 5 or more stations; 622 (25%) at 4 or more stations; and 1439 (57%) at 3 or more stations. Nevertheless, the double-difference algorithm applied to a dataset that was largely comprised of phase pick differences rather than cross-correlation measurements still resulted in improved locations. Figure 9 compares a cross-section of the seismicity located by the traditional method and by modern methods (phase picks plus cross-correlation, and double difference). The latter cross-section more clearly shows a lineation that presumably indicates faulting.

(4) Northern California.

SCHAFF and WALDHAUSER (2005) describe results from an application of cross-correlation methods to process the complete digital seismogram data base for northern California to measure accurate differential times for correlated earthquakes observed at common stations. The waveform database includes about 15 million seismograms from about 210,000 local earthquakes between 1984 and 2003. A total of 26 billion cross correlation measurements were performed on a 32-node (64 processor) Linux cluster. All event pairs with separation distances of 5 km or less were processed at all stations that recorded the pair. A total of about 1.7 billion *P*-wave differential times had cross correlation coefficients (CC) of 0.6 or larger. The *P*-wave differential times are often on the order of a factor of ten to a hundred times more accurate than those obtained from routinely picked phase onsets. 1.2 billion *S*-wave differential times were measured with CC > 0.6, a phase not routinely picked at the Northern California Seismic Network because of the generally weak onset of *S*-phases, often obscured by *P*-wave coda. These results show a surprisingly high degree of waveform similarity for most of the Northern California catalog, which is very encouraging for improving earthquake locations. Overall statistics are that for each of about 200,000 events (95% of the total), waveforms have CC values that are greater than 0.7 for at least four stations with one or more other events. 90% of the events meet this criterion at eight or more stations, and 82% of the events in the catalog cross correlate at twelve or more stations. To illustrate the spatial distribution of correlated events,

Figure 10 shows the percentage of events, within bins of 5 km x 5 km, that have CC > 0.7 for *P*-wave trains with at least one other event at four or more stations. Even tectonically complicated zones exhibit favorable statistics, such as Long Valley Caldera and Geysers Geothermal Field, where mechanisms are quite variable. Apparently, as long as the earthquake density is high enough there is a high probability that at least one other event occurs nearby with a similar focal mechanism, enabling very precise relative locations. To first order, the areas of highly correlated events agree with areas of dense seismicity, suggesting that the closeness of events is the main factor for producing similar waveforms. In a preliminary relocation of 80,000 of these Northern California events, WALDHAUSER et al. (2004b) obtained results exhibiting detailed features, such as streaks of seismicity, very similar to those obtained in previous studies of event subsets such as the Calaveras Fault seismicity.

Discussion and Conclusions

We have assembled and analyzed waveform datasets and phase-pick datasets to assess the capability of modern methods of seismic event location, as compared to traditional methods, for studies of seismicity in China and North America. After making waveform cross-correlation (WCC) measurements within each station archive and combining them with phase-pick differences, the double-difference (DD) relocation algorithm has been applied for each of the study areas. The resulting location estimates are significantly improved over traditional estimates. In some cases, the improvement is very significant and is associated with a high percentage of events that cross correlate. But even without such CC measurements, there is improvement.

Many other authors have carried out related studies. Thus, RUBIN et al. (1999) applied modern methods to relocate about 75% of the events on a creeping section of the San Andreas fault, California. PENG and BEN-ZION (2005) have reported their analysis of about 18,000 aftershocks in the source regions (approximately 80 km by 40 km) of the 1999 Izmit and Duzce

earthquakes in North Anatolia, Turkey. The events had been located quite accurately using traditional methods, and though these authors did not relocate the events they cross-correlated signals at common stations of a ten-station PASSCAL deployment for those events estimated to be not more than 10 km apart. They found that about 60% of the events formed 2000 similar-event clusters (442 of which had five or more events).

For the four regions we examined in this paper, China was studied with far-regional signals and the other three regions were studied with local stations. Yet we found significant differences in the percentage of improved relocations even among the latter three. Specifically we found that about 85% of New Madrid events could be relocated with modern methods (using waveforms cross-correlation), and 95% of Northern California events; but only about 25% of Charlevoix events. It is interesting to find that New Madrid and Northern California have approximately the same station density (about one station per 100 km^2) and distance between stations (about 12 km) whereas Charlevoix is a factor of two smaller in density and a factor of two greater in station distance. Also, we found that primarily seven of the Charlevoix stations provided almost all the useful cross-correlation measurements. The other stations of the network in Eastern Canada are at greater distance, and typically can contribute only phase picks plus just a few correlations to the relocation of Charlevoix events. It is therefore easy to see why a criterion of "four or more" or "five or more" stations is hard to achieve for the events in this part of Eastern Canada: first, because greater epicentral distances are involved to reach four stations, and second because more than half the stations had to record the event with good signal-to-noise ratios and high similarity. For four out of the many more stations in the New Madrid region, good cross-correlation is easier to achieve.

We conclude that the fraction of seismicity amenable to good relocations based on modern methods, ranges from about 10% across some broad regions to as much as 95% in other broad regions (major parts of Northern California). In some cases the percentage is influenced significantly by station density (and clearly this is the case for our study of China using a sparse network). We note that for purposes of monitoring a particular region—whether it be large such

as North America or small such as a particular part of Liaoning Province, China—it is often true that archives of digital seismograms for events in that region are accumulating rapidly and will become more and more capable of supplying events that cross-correlate. Modern methods are most suited to seismically active areas, since these allow cross-correlatable events to accumulate in a shorter time.

The choice of CC threshold has an influence on overall results, in that there is obviously a tradeoff between using as many relative arrival times measured from cross-correlation as possible, but not using measurements of dubious quality. The choice of threshold is best made at a level that improves upon the accuracy of phase picks, and this resulted in thresholds of about 0.7 in our relocation studies using local stations. The DD algorithm eliminates outlier measurements, as long as they are not too numerous.

We note a conceptual inconsistency in that travel-time differences are determined by cross correlating with signals having a finite frequency band, and then measurements are passed to a location algorithm that assumes infinite-frequency ray theory. At event separations comparable to the shortest wavelength and comparable to the spatial scale of medium inhomogeneities, it may be more appropriate to use a location algorithm that explicitly recognizes the finite frequency of the underlying signals. A related point is that we have not explored the effects of material interfaces within the source region, and to the extent that such interfaces can substantially change the take-off directions of signals used for event location (as discussed for example in the case of fault-zone head waves by BEN-ZION and MALIN, 1991), there will be a distorting effect on relative locations, especially for events observed at a limited range of azimuths.

Although aspects of this work can now be regarded as quite routine, it continues to be of importance to acquire practical experience—for example, to provide feedback on questions such as: what station density is required to achieve enough CC measurements to locate the typical event; and what is the effect of archive longevity (how rapidly does the percentage of cross-correlatable events rise, as the data accumulates)? With such practical experience, it may then be

appropriate to consider making modern methods of event location a part of standard operational practice, in support of publication of bulletins of seismicity which would then be enabled to attain much higher quality in seismically active areas where waveforms of neighboring events cross-correlate.

The key to such an approach will be a commitment to long-term station operation, and building up waveform archives that can easily be used for large-scale searches. In practice, it is those network operators and their clients — who have already supported the effort to consistently collect, apply quality control, and build easily-usable archives for seismic waveforms — who are likely to benefit the most from the application of modern methods of event location.

Concerning computing power, we note that for a network consisting of tens of stations and thousands of events, a modern workstation is adequate. But for processing data in the amounts archived for example by the Northern California Earthquake Data Center (approximately a thousand stations, and a quarter of a million events), hardware with the necessary capability has only recently become widely available. The application of modern methods to major (terabyte scale) waveform archives, now being built for different networks and data centers, may require some further applied research at the scale of the full dataset. A test bed that enables access to online archives could be the best facility to acquire such practical experience, prior to operational application of modern methods of seismic event location (waveform cross-correlation and multi-event location algorithms) on a large scale.

Once an archive has been established, and the signals of detected events have been associated and relocated by modern methods, new events can be accurately relocated relative to previously located events. After a period of time, say on the order of a year, all the events associated with the original waveform archive plus a year's worth of additional waveforms can be relocated all together to form a new archive, suitable as the reference set against which to locate the next year's events, and so on.

The context of this work is quite broad, since good locations are commonly the starting point for quantitative seismological studies including tomography and improved knowledge of

Earth structure, seismic hazard analysis, earthquake physics and the interaction between neighboring events in a sequence, and the monitoring of explosions and earthquakes.

Finally we note that waveform cross-correlation can be used not only to measure time differences and relative amplitudes very accurately, but to improve detection of small events, still allowing time differences to be measured even in cases where no phase picks can be made with confidence. See for example, GIBBONS and RINGDAL (2005). If waveform cross-correlation can drop detection thresholds by about one magnitude unit, then approximately ten times more events may be detected, which in turn can lead (via multi-event location algorithms) to significant improvement in the location of all the events, including those of larger magnitude.

Acknowledgements

We greatly appreciate the assistance of network operators associated with the University of Memphis, the Geological Survey of Canada, the Northern California Seismic Network, and the Northern California Earthquake Data Center (NCEDC). In particular we thank David McCormack and Jim Lyons at the Geological Survey of Canada, Ottawa, who helped assemble phase and waveform data from earthquakes in the Charlevoix region; Jer-Ming Chiu of CERI, University of Memphis, and Jiakang Xie at Lamont-Doherty who helped us to acquire phase-pick and waveform data from the PANDA deployment at New Madrid. We also thank Mitch Withers at CERI, University of Memphis for New Madrid seismic network data and Doug Neuhauser for assistance with the NCEDC archive; and Yehuda Ben-Zion, Stephanie Ross and Malcolm Sambridge for constructive comments on the originally submitted ms. This work was supported by the Air Force Research Laboratory under contract F19628-03-C-0129. This is Lamont-Doherty Earth Observatory contribution number 6867.

References

BEN-ZION, Y., and MALIN, P. (1991), *San Andreas fault zone head waves near Parkfield, California*, Science, *251*, 1592–1594.

BONDÁR, I., MYERS, S.C., ENGDAHL, E. R., and BERGMAN, E. A. (2004), *Epicenter accuracy based on seismic network criteria*, Geophys. J. Int., *156*, 483–496.

CHIU, J.M., JOHNSTON, A.J. and YANG, Y.T. (1992), *Imaging the active faults of the central New Madrid seismic zone using PANDA array data*, Seismol. Res. Lett., *63*, 375–393.

DOUGLAS, A. (1967), *Joint epicentre determination*, Nature, *215*, 47–48.

ENGDAHL, E.R., VAN DER HILST, R., and BULAND, R. (1998), *Global teleseismic earthquake relocation with improved travel times and procedures for depth determination*, Bull. Seismol. Soc. Am., *88*, 722–743.

FROHLICH, C. (1979), *An efficient method for joint hypocenter determination for large groups of earthquakes*, Computers and Geosciences, *5*, 3-4, 387–389.

GIBBONS, S.J., and RINGDAL, F. (2005), *The detection of rockbursts at the Barentsburg coal mine, Spitsbergen, using waveform correlation on SPITS array data*, NORSAR Semiannual Technical Summary, 1 July – 31 December 2004, 35–48.

GOT, J.L., FRÉCHET, J., and KLEIN, F.W. (1994), *Deep fault geometry inferred from multiple relative relocation beneath the south flank of Kilauea*, J. Geophys. Res., *99*, 375–393.

HARRIS, D.B. (1991), *A waveform correlation method for identifying quarry explosions*, Bull. Seismol. Soc. Am., *81*, 2395–2418.

HAUKSSON, E., and SHEARER, P. (2005), *Southern California hypocenter relocation with waveform cross-correlation: Part 1. Results using the double-difference method*, Bull. Seismol. Soc. Am., *95*, 896–903.

ISRAELSON, H. (1990), *Correlation of waveforms from closely spaced regional events*, Bull. Seismol. Soc. Am., *80*, 2177–2193.

JORDAN, T.H., and SVERDRUP, K.A. (1981), *Teleseismic location techniques and their application to earthquake clusters in the South-Central Pacific*, Bull. Seismol. Soc. Am., *71*, 1105–1130.

PENG, Z., and BEN-ZION, Y. (2005), *Spatiotemporal variations of crustal anisotropy from similar events in aftershocks of the 1999 M7.4 Izmit and M7.1 Duzce, Turkey, earthquake sequences*, Geophys. J. Int., *160*, 1027–1043.

POUPINET, G., ELLSWORTH, W.L. and FRÉCHET, J. (1984), *Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California*, J. Geophys. Res., *89*, 5719–5731.

RICHARDS, P.G., ARMBRUSTER, J., BURLACU, V., CORMIER, V.F., FISK, M.D., KHALTURIN, V.I., KIM, W.-Y., MOROZOV, I.B., MOROZOVA, E.A., SAIKIA, C.K., SCHAFF, D., STROUJKOVA, A. and WALDHAUSER, F. (2003), *Seismic Location Calibration for 30 International Monitoring System Stations in Eastern Asia: Final Results*, Proceedings, DOD/DOE Seismic Research Review, Tucson.

RICHARDS-DINGER, K.B., and SHEARER, P.M. (2000), *Earthquake locations in Southern California obtained using source-specific station terms*, J. Geophys. Res., *105*, 10,939–10,960.

ROWE, C.A., ASTER, R.C., PHILLIPS, W.S., JONES, R.H., BORCHERS, B., and FEHLER, M.C. (2002), *Using automated, high-precision repicking to improve delineation of microseismic structures at the Soultz Geothermal Reservoir*, Pure Appl. Geophys., *159*, 563–596.

RUBIN, A.M., GILLARD, D. and GOT, J.L. (1999), *Streaks of microearthquakes along creeping faults*, Nature, *400*, 635–641.

SCHAFF, D.P., BOKELMANN, G.H.R., BEROZA, G.C., WALDHAUSER, F. and ELLSWORTH, W.L. (2002), *High-resolution image of Calaveras Fault seismicity*, J. Geophys. Res., *107*, 2186, doi: 1029/2001JB000633.

SCHAFF, D.P., and RICHARDS, P.G. (2004a), *Repeating seismic events in China*, Science, *303*, 1176–1178.

SCHAFF, D.P., and RICHARDS, P.G. (2004b), *Lg-wave cross correlation and double-difference location: application to the 1999 Xiuyan, China, sequence*, Bull. Seismol. Soc. Am., *94*, 867–879.

SCHAFF, D.P., BOKELMANN, G.H.R., BEROZA, G.C. and ELLSWORTH, W.L. (2004), *Optimizing correlation techniques for improved earthquake location*, Bull. Seismol. Soc. Am., *94*, 705–721.

SCHAFF, D.P., and WALDHAUSER, F. (2005), *Waveform cross correlation based differential travel time measurements at the Northern California Seismic Network*, Bull. Seismol. Soc. Am., in press.

SHEARER, P. (1997), *Improving local earthquake locations using the L1 norm and waveform cross correlation: application to the Whittier Narrows, California, aftershock sequence*, J. Geophys. Res., *102*, 8269–8283.

SHEARER, P., HAUKSSON, E. and LIN, G. (2005), *Southern California hypocenter relocation with waveform cross-correlation: Part 2. Results using source-specific station terms and cluster analysis*, Bull. Seismol. Soc. Am., *95*, 904–915.

THURBER, C., TRABANT, C., HASLINGER, F. and HARTOG, R. (2001), *Nuclear explosion locations at the Balapan, Kazakhstan, nuclear test site: the effects of high-precision arrival times and three-dimensional structure*, Phys. Earth Plan. Int., *123*, 283–301.

WALDHAUSER, F. and ELLSWORTH, W.L. (2000), *A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California*, Bull. Seismol. Soc. Am., *90*, 1353–1368.

WALDHAUSER, F. (2001), *HypoDD: a program to compute double-difference hypocenter locations*, U.S.G.S. open-file report, 01–113, Menlo Park, California.

WALDHAUSER, F., ELLSWORTH, W.L. and COLE, A. (1999), *Slip-parallel seismic lineations along the Northern Hayward Fault, California*, Geophys. Res. Lett., *26*, 3525–3528.

WALDHAUSER, F., SCHAFF, D.P., RICHARDS, P.G. and KIM, W.-Y. (2004a), *Lop Nor revisited: underground nuclear explosion locations, 1976–1996, from double-difference analysis of regional and teleseismic data*, Bull. Seismol. Soc. Am., *94*, 1879–1889.

WALDHAUSER, F., SCHAFF, D.P., KIM, W.-Y., and RICHARDS, P.G. (2004b), *Improved characterization of seismicity and fault structure by wide area event relocation*, EOS, Trans. Am. Geophys. U., *85*, Fall Meet. Suppl., Abstract S53C–07.

WITHERS, M., ASTER, R., and YOUNG, C. (1999), *An automated local and regional seismic event detection and location system using waveform correlation*, Bull. Seismol. Soc. Am., *86*, 657–669.

XIE, J., LIU, Z., CONG, L., HERRMANN, R.B. and CHIU, J.M. (1997), *Rupture properties of clustered microearthquakes near intersecting intraplate faults of the New Madrid Seismic Zone: Implications on fault weakening*, J. Geophys. Res., *102*, 8,187–8,202.

XIE, J. (2001), *Rupture characteristics of clustered microearthquakes and variations in fault properties in the New Madrid Seismic Zone*, J. Geophys. Res., *106*, 26,495–26,509.

cross-correlation value (CC)	**% of events in ABCE**
0.9	3
0.8	9
0.7	16
0.6	23
0.5	30
0.4	38
0.3	52
0.2	69
0.1	76

Table 1. The percentage of events in the Annual Bulletin of Chinese Earthquakes (ABCE) is shown for different values of the cross-correlation value. The effect of relaxing the stringent criterion ($CC \geq 0.8$) is substantial. (The analysis was done for 14,000 ABCE events, 130,000 seismograms, 1.2 million CC.)

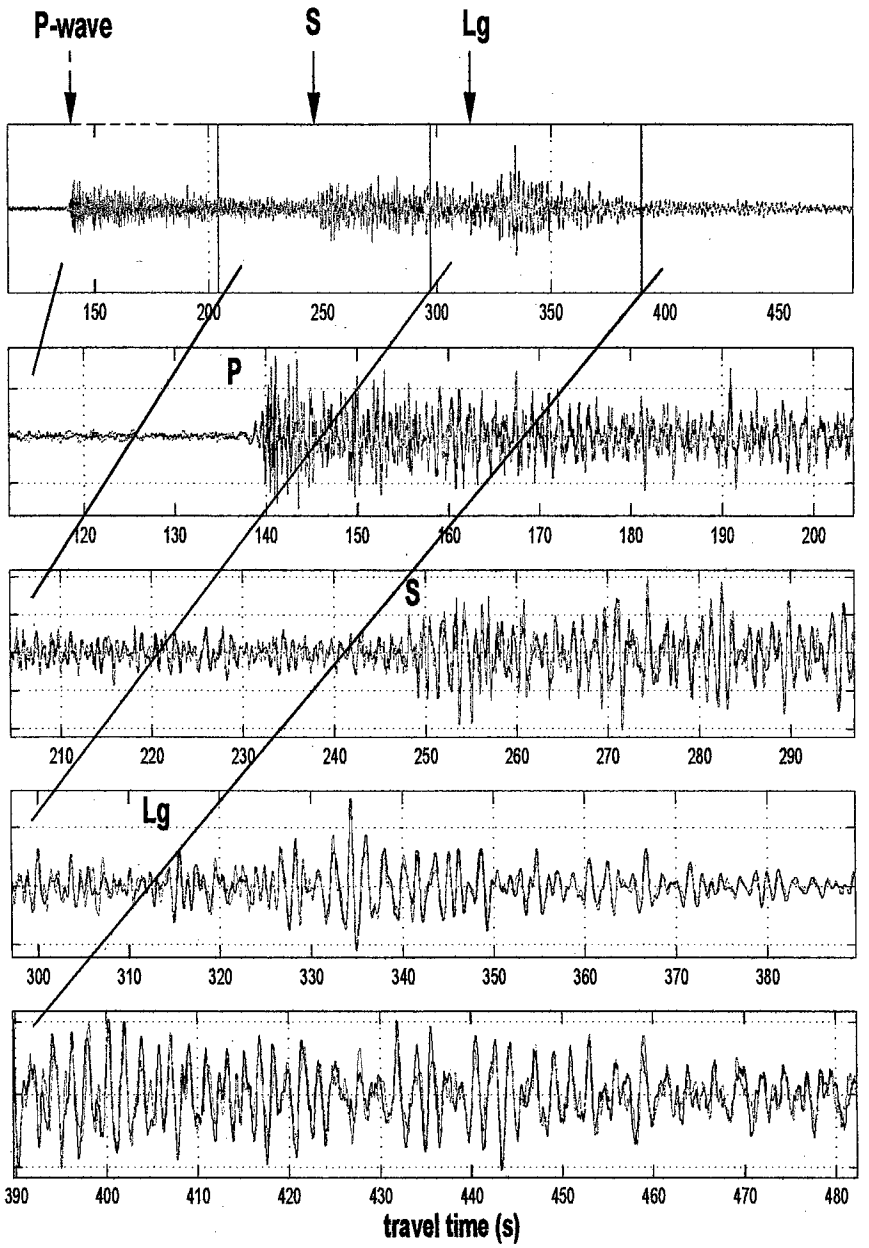

Figure 1. A pair of similar events in China filtered from 0.5 to 5 Hz and superimposed in gray and black. Vertical axes are normalized to unit amplitude. Lower subpanels are enlargements of the white and gray segments. The predicted *P*-wave arrives at 143 s, the *S*-wave at 256 s, and the *Lg*-wave at 315 s. It is apparent that these waveforms are very similar. The quality of this waveform doublet is typical of the events described in more detail by SCHAFF and RICHARDS (2004ab), who argue that such events must be not more than about 1 km apart.

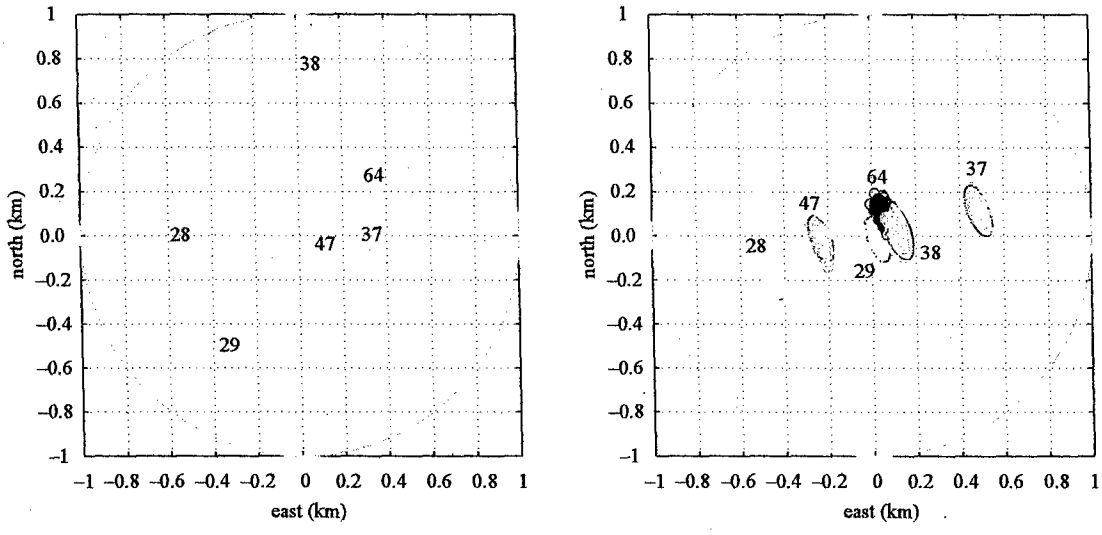

Figure 2. Comparison of double-difference relative locations for a subset of events in the sequence for a local/regional network (left) using only *P*-wave phase picks recorded at several hundred stations and for a sparse regional network (archived at IRIS) using *Lg* cross-correlation measurements (right). Event numbers are for identification. The RMS travel time residuals are about 1 sec for the *P*-waves and 0.02 sec for *Lg*. 95% confidence formal error ellipses and bootstrap errors (shaded small circles) are in good agreement (right). The epicenter in each case is taken as the centroid of the cluster. For further details see SCHAFF and RICHARDS (2004b).

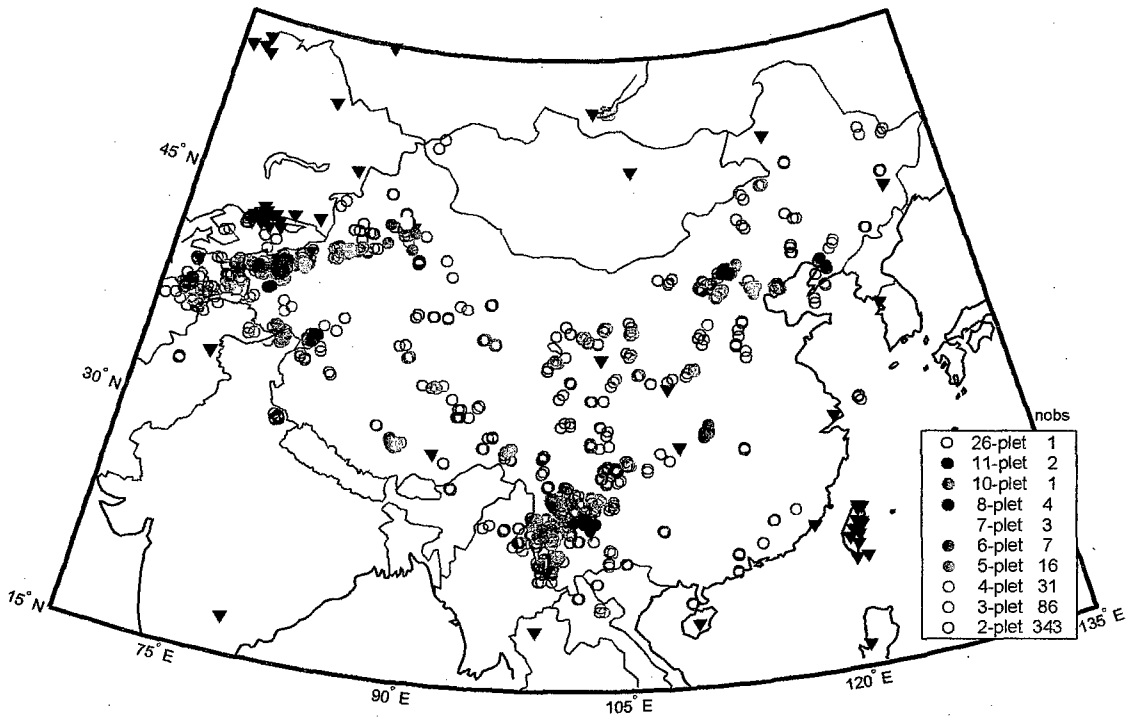

Figure 3. 1301 events (9% of the Annual Bulletin of Chinese Earthquakes -- ABCE), whose seismograms satisfy the criterion of cross correlation coefficients greater than or equal to 0.8 with seismograms from at least one other earthquake, for long windows from 5 seconds before the *P*-wave to 40 sec after the *Lg*-wave on waveforms that are filtered from 0.5 to 5 Hz (as shown in Fig. 1). There are 494 multiplets here, and the inset shows the number of multiplets containing between 2 and 26 events. Recording stations archived by the IRIS Consortium are denoted with solid black triangles. Events are plotted at their ABCE absolute locations. For further details see SCHAFF and RICHARDS (2004a).

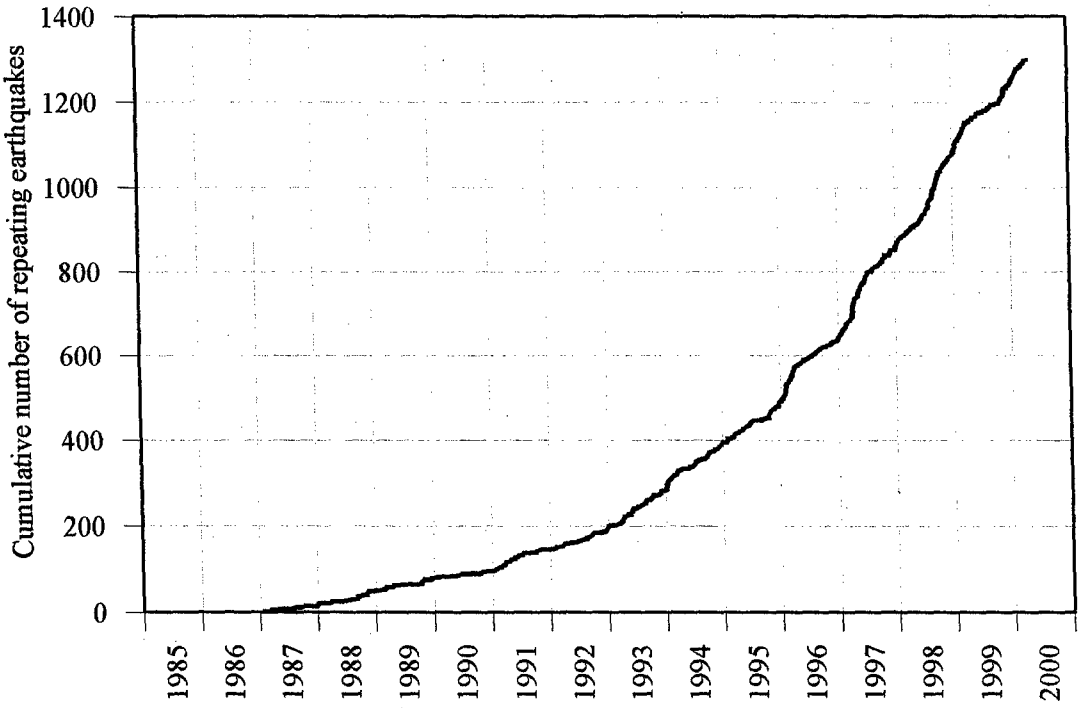

Figure 4. Cumulative number of events vs. time for the 1301 repeating events in China shown in Figure 3 for the period from 1985 to early 2000. Network coverage is more sparse for earlier years, resulting in underestimation of the actual percentage of repeating events over the entire period. If only the 908 repeating events later than 1994 are considered (out of the total of about 6400 events reported in the Annual Bulletin of Chinese Earthquakes), then about 14% of the events satisfy the stringent criterion of cross-correlation not less than 0.8 for a window from the *P*-arrival until 40 s into the *Lg* signal.

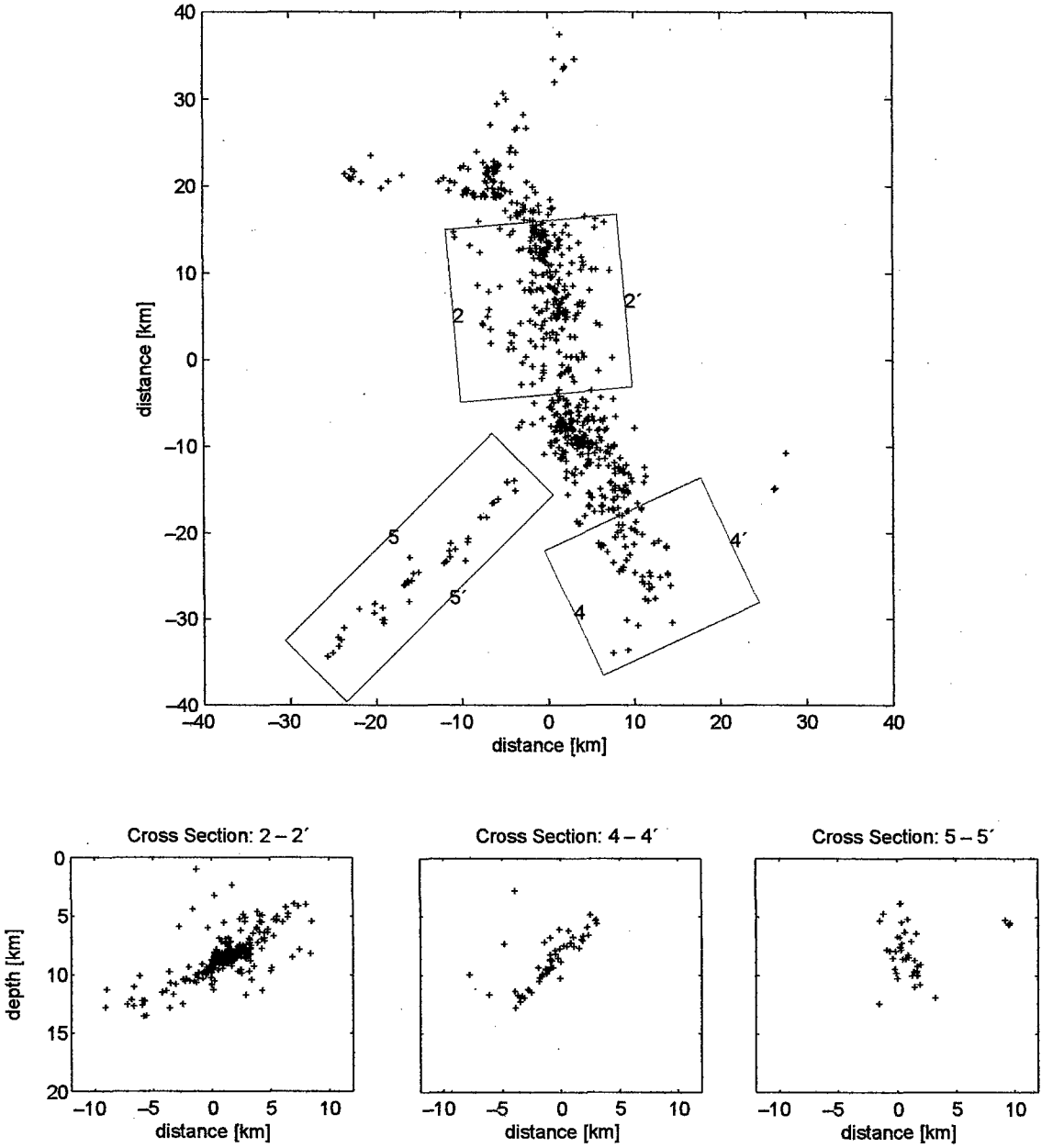

Figure 5. Map view and three cross sections for 783 relocated seismic events in the New Madrid region of the central United States from October 1989 to August 1992, using 42 stations of a PANDA network (see CHIU et al., 1992). This application of the double-difference algorithm was based upon 58,807 phase pick pairs and 80,697 cross-correlations (CC ≥ 0.7) derived from *P*-waves, plus 66,581 phase pick pairs and 81,750 cross-correlations (CC ≥ 0.7) derived from *S*-waves. Cross sections show clear evidence for a westward-dipping fault plane, the dip increasing from about 30° (section 2–2′) to 45° further south (section 4–4′).

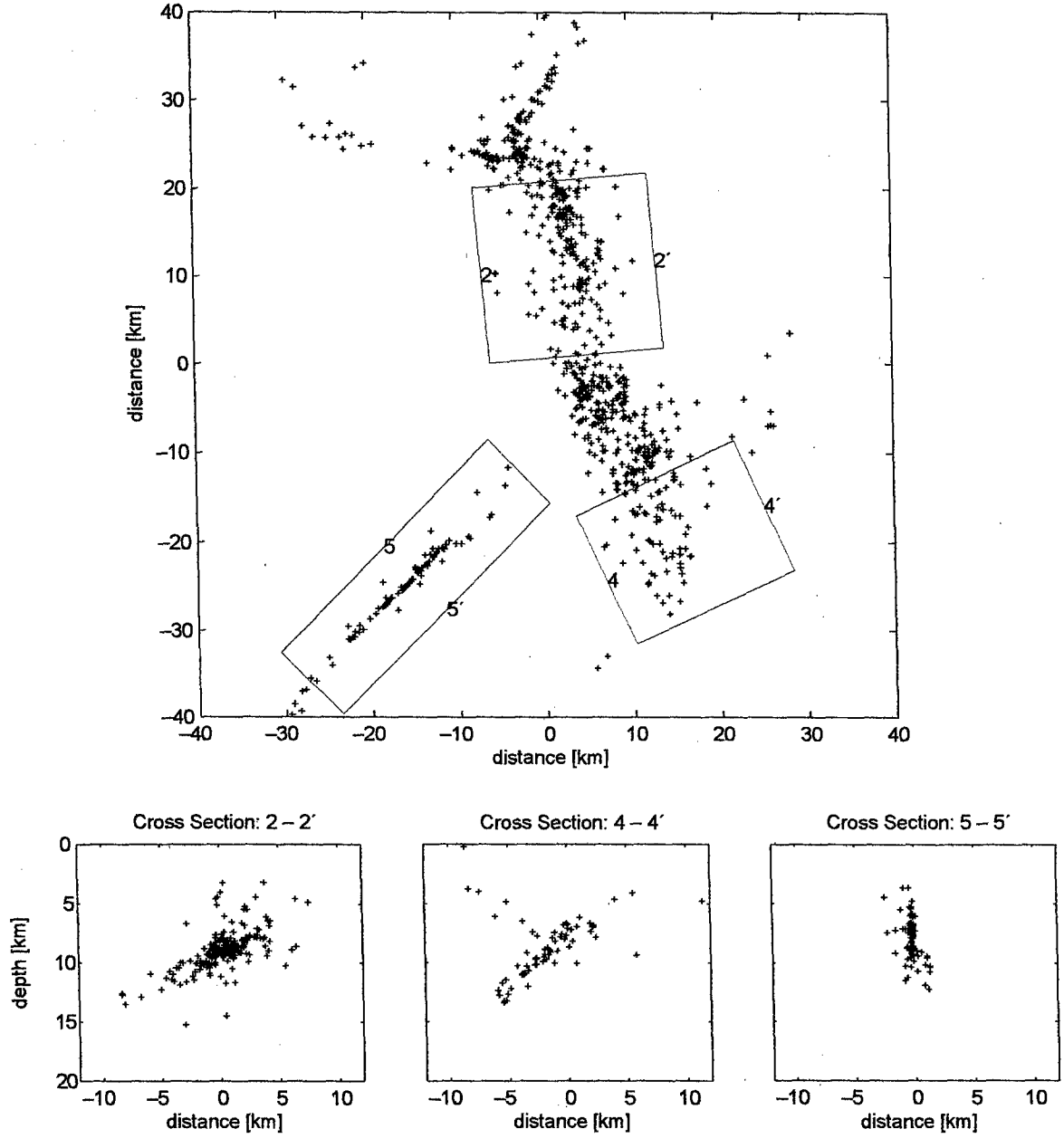

Figure 6. Map view and three cross sections for 594 relocated seismic events in the New Madrid region from January 2000 to October 2003, using 85 stations operated by the University of Memphis. This application of the double-difference algorithm was based upon 202,249 phase pick pairs and 49,660 cross-correlations (CC ≥ 0.7) derived from *P*-waves, plus 192,277 phase pick pairs and 93,009 cross-correlations (CC ≥ 0.7) derived from *S*-waves.

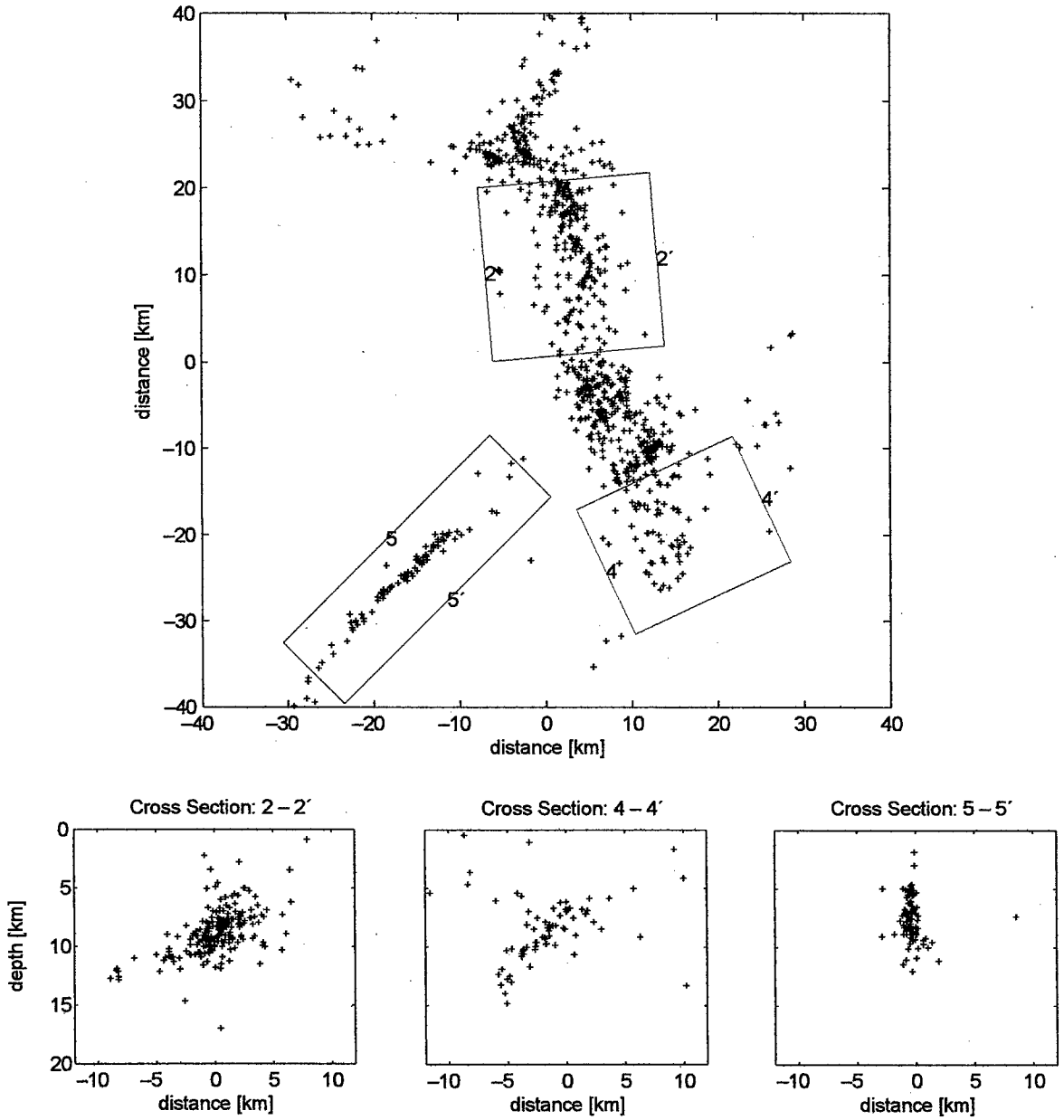

Figure 7. The traditionally-obtained event locations (catalog locations, based on phase picks and analysis of events one-at-a-time), in map view and for three cross sections, for the station set and time period of Figure 6.

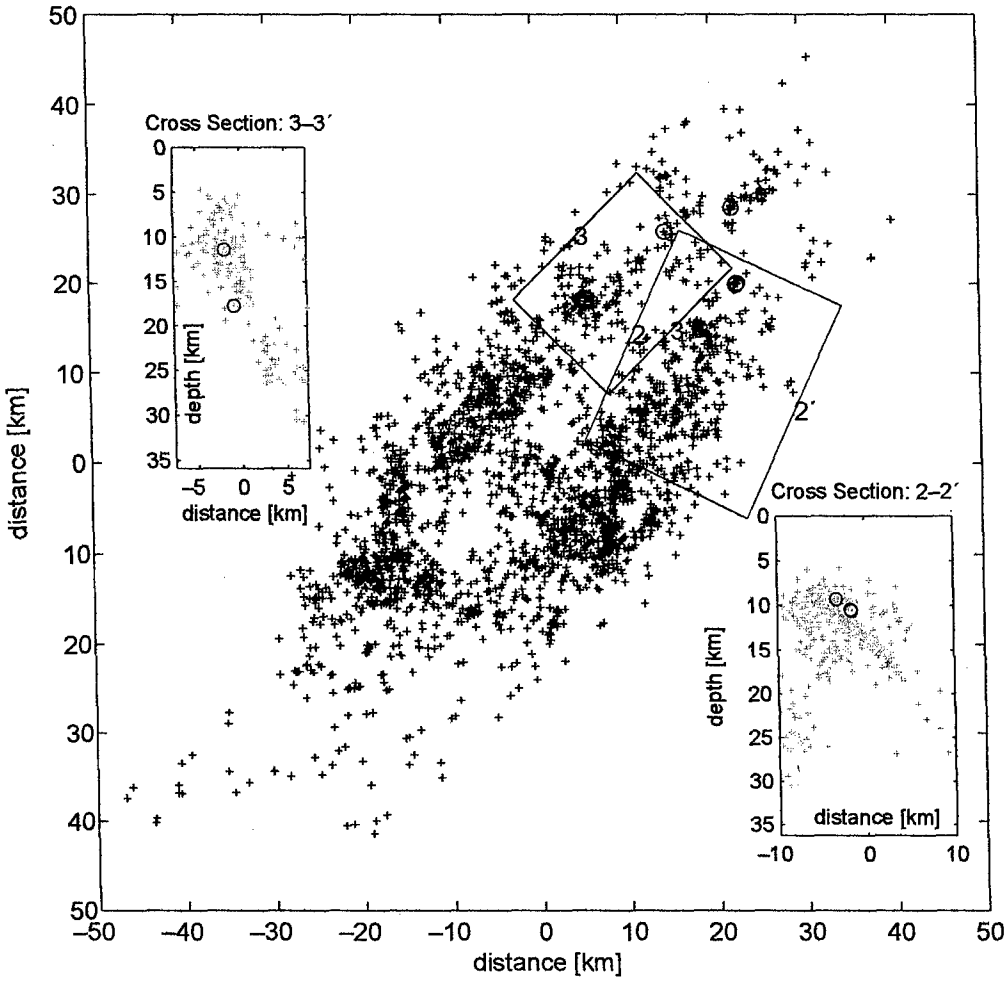

Figure 8. Relocation of 2272 events in the Charlevoix region of Eastern Canada from January 1988 to December 2003, using 46 stations operated by the Geological Survey of Canada. This application of the double-difference algorithm was based upon 495,347 phase pick pairs and 127,600 cross-correlations (CC ≥ 0.7) derived from *P*-waves, plus 567,823 phase pick pairs and 153,510 cross-correlations (CC ≥ 0.7) derived from *S*-waves. Circled events have magnitude ≥ 4.

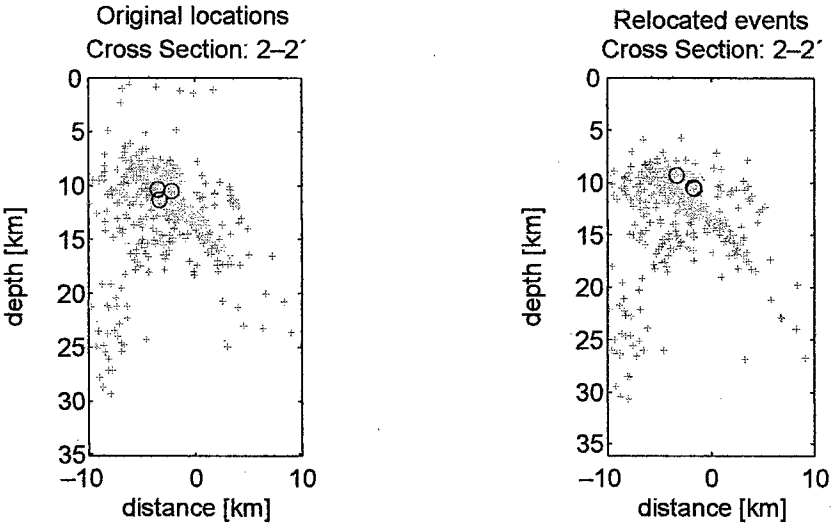

Figure 9. For the Charlevoix cross-section 2–2′ of Figure 8, here are shown the locations obtained by traditional methods (on the left) and the relocations using phase picks plus a relatively small number of relative arrival times measured via cross-correlation, and relocating events all together. The relocated events more clearly delineate a linear structure, dipping about 45° to the southeast.

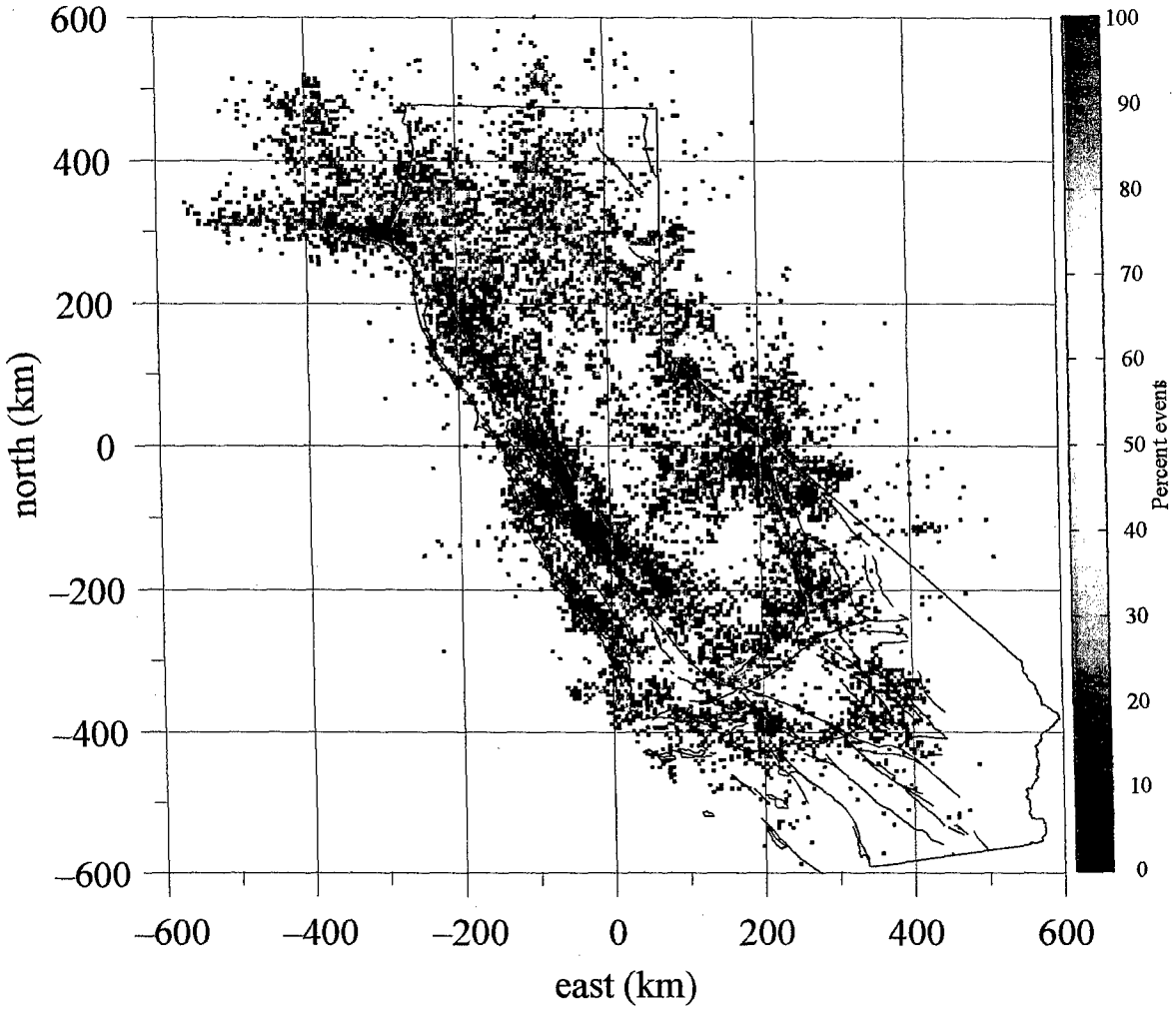

Figure 10. There are more than 200,000 events in the Northern California catalog since 1984 for which waveforms cross correlate (at CC > 0.7) with at least one other event at four or more stations. Here is shown the percentage of events in 5 km by 5 km bins that meet this criterion. For further details see SCHAFF and WALDHAUSER (2005).